

SuperFect (Qiagen) and pulse-labeled for 3 hours with [<sup>35</sup>S]methionine and [<sup>35</sup>C]cysteine. Both epitope-tagged proteins co-migrate when 20 microliters of 15-fold concentrated serum-free conditioned medium were electrophoresed on a polyacrylamide gel (Novex) in sodium dodecyl sulfate sample buffer (SDS-PAGE). The VEGF-E-IgG expression plasmid was constructed by cloning the ORF in front of the human Fc (IgG) sequence.

The VEGF-E-IgG plasmid was co-transfected with Baculogold Baculovirus DNA (Pharmlngen) using Lipofectin (GibcoBRL) into 10<sup>5</sup> Sf9 cells grown in Hink's TNM-FH medium (JRH Biosciences) supplemented with 10% fetal bovine serum. Cells were incubated for 5 days at 28°C. The supernatant was harvested and subsequently used for the first viral amplification by infecting Sf9 cells at an approximate multiplicity of infection (MOI) of 10. Cells were incubated for 3 days, then supernatant harvested, and expression of the recombinant plasmid determined by binding of 1 ml of supernatant to 30 µl of Protein-A Sepharose CL-4B beads (Pharmacia) followed by subsequent SDS-PAGE analysis. The first amplification supernatant was used to infect a 500 ml spinner culture of Sf9 cells grown in ESF-921 medium (Expression Systems LLC) at an approximate MOI of 0.1. Cells were treated as above, except harvested supernatant was sterile filtered. Specific protein was purified by binding to Protein-A Sepharose 4 Fast Flow (Pharmacia) column.

#### EXAMPLE 86: Northern Blot Analyses for PRO200

Blots of human poly(A)<sup>+</sup> RNA from multiple adult and fetal tissues and tumor cell lines were obtained from Clontech (Palo Alto, CA). Hybridization was carried out using <sup>32</sup>P-labeled probes containing the entire coding region and washed in 0.1 x SSC, 0.1% SDS at 63°C.

VEGF-E mRNA was detectable in fetal lung, kidney, brain, liver and adult heart, placenta, liver, skeletal muscle, kidney, and pancreas. VEGF-E mRNA was also found in A549 lung adenocarcinoma and HeLa cervical adenocarcinoma cell lines.

#### EXAMPLE 87: In Situ Hybridization of Human Fetal Tissue Sections for PRO200

Formalin-fixed, paraffin-embedded human fetal brain, liver, lower limb, small intestine, thyroid, lymph node, thymus, stomach, trachea, skin, spleen, spinal cord, adrenal, placenta, cord, and adult liver, pancreas, lung, spleen, lymph node, adrenal, heart, aorta, and skin were sectioned, deparaffinized, deproteinized in proteinase K (20 µg/ml) for 15 minutes at 37°C, and further processed for in situ hybridization as described by Lu LH and Gillett NA (Cell Vision 1:169-176, 1994). A [ $\alpha$ -<sup>33</sup>P]UTP-labeled antisense riboprobe was generated from a PCR product of 980 bp (primers GCGGAATCCAACTGAGTAG and GCGGCTATCCTCTGTGCTC, SEQ ID NOS: 493 and 494, respectively). The slides were dipped in Kodak NTB2 nuclear track emulsion and exposed for 4 weeks.

VEGF-E mRNA expression included localization at the growth plate region and embracing fetal myocytes.

#### EXAMPLE 88: Myocyte Hypertrophy Assay for PRO200

Myocytes from neonatal Harlan Sprague Dawley rat heart ventricle (23 days gestation) were plated in duplicate at 75000 cells/ml in a 96-well plate. Cells were treated for 48h with 2000, 200, 20, or 2 ng/ml VEGF-E-IgG. Myocytes were stained with crystal violet to visualize morphology and scored on a scale of 3 to

7, 3 being nonstimulated and 7 being full-blown hypertrophy.

2000 ng/ ml and 200 ng/ ml VEGF-E caused hypertrophy, scored as a 5.

EXAMPLE 89: Cell Proliferation Assay for PRO200

Mouse embryonic fibroblast C3H10T1/2 cells (ATCC) were grown in 50:50 Ham's F-12: low glucose DMEM medium containing 10% fetal calf serum (FCS). Cells were plated in duplicate in a 24-well plate at 1000, 2000, and 4000 cells/well. After 48 hours, cells were switched to medium containing 2% FCS and were incubated for 72 hours with 200, 800, or 2000 ng/ml VEGF-E or no growth factor added.

Approximately 1.5 fold greater number of cells were measured in the presence of 200 ng/ml VEGF-E as in its absence, at all three cell densities.

EXAMPLE 90: Endothelial Cell Survival Assay for PRO200

Human umbilical vein endothelial cells (HUVEC, Cell Systems) were maintained in Complete Media (Cell Systems) and plated in triplicate in serum-free medium (Basic Media from Cell Systems containing 0.1% BSA) at 20,000 cells/well of a 48-well plate. Cells were incubated for 5 days with 200 or 400 ng/ml VEGF-E-IgG, 100 ng/ml VEGF, 20 ng/ml basic FGF, or no addition.

Survival was 2-3 times greater with VEGF-E as compared to lack of growth factor addition. VEGF and basic FGF were included as positive controls.

EXAMPLE 91: Isolation of cDNA Clones Encoding Human PRO285

A proprietary expressed sequence tag (EST) DNA database (LIFESEQ™, Incyte Pharmaceuticals, Palo Alto, CA) was searched and an EST (#2243209) was identified which showed homology to the *Drosophila* Toll protein.

Based on the EST, a pair of PCR primers (forward and reverse):

TAAAGACCCAGCTGTGACCG (SEQ ID NO:499)

ATCCATGAGCCTCTGATGGG (SEQ ID NO: 500), and

a probe:

ATTTATGTCTCGAGGAAAGGGACTGGTTACCAAGGCAGCCAGTTC (SEQ ID NO: 501)

were synthesized.

mRNA for construction of the cDNA libraries was isolated from human placenta tissue. The cDNA libraries used to isolate the cDNA clones were constructed by standard methods using commercially available reagents such as those from Invitrogen, San Diego, CA (Fast Track 2). The cDNA was primed with oligo dT containing a NotI site, linked with blunt to SalI hemikinased adaptors, cleaved with NotI, sized appropriately by gel electrophoresis, and cloned in a defined orientation into the cloning vector pCR2.1 (Invitrogen, Inc.) using reagents and protocols from Life Technologies, Gaithersburg, MD (Super Script Plasmid System). The double stranded cDNA was sized to greater than 1000 bp and the cDNA was cloned into BamHI/NotI cleaved vector. pCR2.1 is a commercially available plasmid, designed for easy cloning of PCR fragments, that carries AmpR and KanR genes for selection, and LacZ gene for blue-white selection.

In order to screen several libraries for a source of a full-length clone, DNA from the libraries was screened by PCR amplification with the PCR primer pair identified above. A positive library was then used to isolate clones encoding the PRO285 gene using the probe oligonucleotide and one of the PCR primers.

5 A cDNA clone was sequenced in entirety. The entire nucleotide sequence of DNA40021-1154 (encoding PRO285) is shown in Figure 208 (SEQ ID NO:495). Clone DNA40021-1154 contains a single open reading frame with an apparent translational initiation site at nucleotide positions 61-63 (Figure 208). The predicted polypeptide precursor is 1049 amino acids long, including a putative signal peptide at amino acid positions 1-29, a putative transmembrane domain between amino acid positions 837-860, and a leucine zipper pattern at amino acid positions 132-153 and 704-725, respectively. It is noted that the indicated boundaries are approximate, and the actual limits of the indicated regions might differ by a few amino acids. Clone DNA40021-1154 has been deposited with ATCC (designation: DNA40021-1154) and is assigned ATCC deposit no.209389.

Based on a BLAST and FastA sequence alignment analysis (using the ALIGN computer program) of the full-length sequence is a human analogue of the *Drosophila* Toll protein, and is homologous to the following human Toll proteins: Toll1 (DNAX# HSU88540-1, which is identical with the random sequenced full-length cDNA #HUMRSC786-1); Toll2 (DNAX# HSU88878-1); Toll3 (DNAX# HSU88879-1); and Toll4 (DNAX# HSU88880-1).

#### EXAMPLE 92: Isolation of cDNA Clones Encoding Human PRO286

A proprietary expressed sequence tag (EST) DNA database (LIFESEQ™, Incyte Pharmaceuticals, Palo Alto, CA) was searched and an EST (#694401) was identified which showed homology to the *Drosophila* Toll protein.

Based on the EST, a pair of PCR primers (forward and reverse):

GCCGAGACAAAACGTTCTCC (SEQ ID NO:502)

25 CATCCATGTTCTCATCCATTAGCC (SEQ ID NO: 503), and

a probe:

TCGACAACCTCATGCAGAGCATCAACCAAGCAAGAAAACAGTATT (SEQ ID NO: 504)

were synthesized.

30 mRNA for construction of the cDNA libraries was isolated from human placenta tissue. This RNA was used to generate an oligo dT primed cDNA library in the vector pRK5D using reagents and protocols from Life Technologies, Gaithersburg, MD (Super Script Plasmid System). pRK5D is a cloning vector that has an sp6 transcription initiation site followed by an SfiI restriction enzyme site preceding the XhoI/NotI cDNA cloning sites. The cDNA was primed with oligo dT containing a NotI site, linked with blunt to SalI hemikinased adaptors, cleaved with NotI, sized to greater than 1000 bp appropriately by gel electrophoresis, and cloned in

35 a defined orientation into XhoI/NotI-cleaved pRK5D.

In order to screen several libraries for a source of a full-length clone, DNA from the libraries was screened by PCR amplification with the PCR primer pair identified above. A positive library was then used to isolate clones encoding the PRO286 gene using the probe oligonucleotide identified above and one of the PCR

primers.

A cDNA clone was sequenced in entirety. The entire nucleotide sequence of DNA42663-1154 (encoding PRO286) is shown in Figure 210 (SEQ ID NO:497). Clone DNA42663-1154 contains a single open reading frame with an apparent translational initiation site at nucleotide positions 57-59 (Figure 211). The predicted polypeptide precursor is 1041 amino acids long, including a putative signal peptide at amino acid positions 1-26, a potential transmembrane domain at amino acid positions 826-848, and leucine zipper patterns at amino acids 130-151, 206-227, 662-684, 669-690 and 693-614, respectively. It is noted that the indicated boundaries are approximate, and the actual limits of the indicated regions might differ by a few amino acids. Clone DNA42663-1154 has been deposited with ATCC (designation: DNA42663-1154) and is assigned ATCC deposit no. 209386.

Based on a BLAST and FastA sequence alignment analysis (using the ALIGN computer program) of the full-length sequence of PRO286, it is a human analogue of the *Drosophila* Toll protein, and is homologous to the following human Toll proteins: Toll1 (DNAX# HSU88540-1, which is identical with the random sequenced full-length cDNA #HUMRSC786-1); Toll2 (DNAX# HSU88878-1); Toll3 (DNAX# HSU88879-1); and Toll4 (DNAX# HSU88880-1).

#### EXAMPLE 93: NF- $\kappa$ B Assay for PRO285 and PRO286

As the Toll proteins signal through the NF- $\kappa$ B pathway, their biological activity can be tested in an NF- $\kappa$ B assay. In this assay Jurkat cells are transiently transfected using Lipofectamine reagent (Gibco BRL) according to the manufacturer's instructions. 1  $\mu$ g pB2XLuc plasmid, containing NF- $\kappa$ B-driven luciferase gene, is cotransfected with 1  $\mu$ g pSR $\alpha$ N expression vector with or without the insert encoding PRO285 or PRO286. For a positive control, cells are treated with PMA (phorbol myristyl acetate; 20 ng/ml) and PHA (phytohaemagglutinin, 2  $\mu$ g/ml) for three to four hours. Cells are lysed 2 or 3 days later for measurement of luciferase activity using reagents from Promega.

#### EXAMPLE 94: Isolation of cDNA Clones Encoding Human PRO213-1, PRO1330 and PRO1449

A consensus DNA sequence was assembled relative to other EST sequences using phrap as described in Example 1 above. This consensus sequence is herein designated DNA28735. Based on the DNA28735 consensus sequence, oligonucleotides were synthesized: 1) to identify by PCR a cDNA library that contained the sequence of interest, and 2) for use as probes to isolate a clone of the full-length coding sequence for PRO213-1, PRO1330 and/or PRO1449. A pair of PCR primers (forward and reverse) were synthesized:

forward PCR primer 5'-TGGAGCAGCAATATGCCAGCC-3' (SEQ ID NO:511)

reverse PCR primer 5'-TTTCCACTCCTGTCGGGTTGG-3' (SEQ ID NO:512)

Additionally, a synthetic oligonucleotide hybridization probe was constructed from the consensus DNA28735 sequence which had the following nucleotide sequence:

hybridization probe

5'-GGTGACACTTGCCAGTCAGATGTGGATGAATGCAGTGCTAGGAGGG-3' (SEQ ID NO:513)

In order to screen several libraries for a source of a full-length clone, DNA from the libraries was screened by PCR amplification with the PCR primer pair identified above. A positive library was then used to isolate clones encoding the PRO213-1, PRO1330 and/or PRO1449 gene using the probe oligonucleotide and one of the PCR primers. RNA for construction of the cDNA libraries was isolated from human fetal lung tissue.

DNA sequencing of the clones isolated as described above gave the full-length DNA sequence encoding PRO213-1, PRO1330 and/or PRO1449 [DNA30943-1-1163-1 (SEQ ID NO:505), DNA64907-1163-1 (SEQ ID NO:507) and DNA64908-1163-1 (SEQ ID NO:509), respectively].

The entire nucleotide sequences corresponding to DNA30943-1-1163-1 (SEQ ID NO:505), DNA64907-1163-1 (SEQ ID NO:507) and DNA64908-1163-1 (SEQ ID NO:509), respectively. DNA30943-1163, DNA64907-1163-1 and DNA64908-1163-1 contain a single open reading frame with an apparent translational initiation site at nucleotide positions 336-338, 488-490 and 326-328, respectively, and ending at the stop codon at nucleotide positions 1221-1223, 1307-1309 and 1145-1147, respectively (Figures 212, 214 and 216). The predicted polypeptide precursor is 295, 273 and 273 amino acids long, respectively (Figures 213, 215 and 217). DNA30943-1-1163-1, DNA64907-1163-1 and DNA64908-1163-1 have been deposited with ATCC and are assigned ATCC deposit no. 209791, 203242 and 203243, respectively.

Analysis of the amino acid sequence of the full-length PRO213-1 polypeptide suggests that a portion of it possess significant homology to the human growth arrest-specific gene 6 protein. More specifically, an analysis of the Dayhoff database (version 35.45 SwissProt 35) evidenced significant homology between the PRO213 amino acid sequence and the following Dayhoff sequences, HSMHC3W5A\_6 and B48089.

Additional analysis of the amino acid sequence of the full-length PRO1330 and PRO1449 polypeptide indicates significant identity with notch4. More specifically, an analysis of the Dayhoff database (version 35.130 SwissProt 35) evidenced significant identity between PRO1330 and the following Dayhoff sequences, D86566\_1 and NEL\_HUMAN.

#### EXAMPLE 95: Isolation of cDNA Clones Encoding Human PRO298

A cDNA isolated in the amylase screen described in Example 2 above is herein designated DNA26832 (Figure 220; SEQ ID NO:516). The sequence of DNA26832 was then used to search expressed sequence tag (EST) databases. The EST databases included public EST databases (e.g., GenBank) and a proprietary EST database (LIFESEQ™, Incyte Pharmaceuticals, Palo Alto, CA). The search was performed using the computer program BLAST or BLAST2 (Altschul et al., *Methods in Enzymology* 266: 469-480 [1996]). Those comparisons resulting in a BLAST score of 70 (or in some cases 90) or greater that did not encode proteins were clustered and assembled into consensus DNA sequences with the program "phrap" (Phil Green, University of Washington, Seattle, Washington; <http://bozeman.mbt.washington.edu/phrap.docs/phrap.html>).

A consensus DNA sequence was assembled relative to other EST sequences using phrap. A consensus sequence was determined, which was then extended using repeated cycles of BLAST and phrap to extend the consensus sequence as far as possible using the sources of EST sequences discussed above. The extended assembly sequence was designated DNA35861. Based on the DNA35861 consensus sequence, oligonucleotides were synthesized: 1) to identify by PCR a cDNA library that contained the sequence of interest, and 2) for use as probes to isolate a clone of the full-length coding sequence of PRO298. Forward and reverse primers

generally range from 20 to 30 nucleotides and are often designed to give a PCR product of about 100-1000 bp in length. The probe sequence is typically 40-55 bp in length. In some cases, additional oligonucleotides are synthesized when the consensus sequence is greater than about 1-1.5kbp. In order to screen several libraries for a full-length clone, DNA from the libraries was screened by PCR amplification, as per Ausubel et al., Current Protocols in Molecular Biology, with the PCR primer pair. A positive library was used to isolate clones

encoding the gene of interest using the probe oligonucleotide and one of the primer pairs.

PCR primers (forward and reverse) and a hybridization probe were synthesized:

forward PCR primer 1 CAACGTGATTTCAAAGCTGGGCTC (SEQ ID NO:517)

forward PCR primer 2 GCCTCGTATCAAGAAATTCC (SEQ ID NO:518)

forward PCR primer 3 AGTGGAAGTCGACCTCCC (SEQ ID NO:519)

reverse PCR primer 1 CTCACCTGAAATCTCTCATAGCCC (SEQ ID NO:520)

hybridization probe 1 CGCAAAACCCATTTTGGGAGCAGGAATTCCAATCATGTCTGTGATGGTGG (SEQ ID NO:521)

In order to screen several libraries for a source of a full-length clone, DNA from the libraries was screened by PCR amplification with the PCR primer pair identified above. A positive library was then used to isolate clones encoding the PRO298 gene using the probe oligonucleotide and one of the PCR primers.

RNA for construction of the cDNA libraries was isolated from human fetal lung tissue (LIB25). The cDNA libraries used to isolate the cDNA clones were constructed by standard methods using commercially available reagents such as those from Invitrogen, San Diego, CA. The cDNA was primed with oligo dT containing a NotI site, linked with blunt to SalI hemikinased adaptors, cleaved with NotI, sized appropriately by gel electrophoresis, and cloned in a defined orientation into a suitable cloning vector (such as pRKB or pRKD; pRK5B is a precursor of pRK5D that does not contain the SfiI site; see, Holmes et al., Science, 253:1278-1280 (1991)) in the unique XhoI and NotI sites.

DNA sequencing of the clones isolated as described above gave the full-length DNA sequence for PRO298 (herein designated UNQ261 [DNA39975-1210]) (SEQ ID NO:514), and the derived protein sequence for PRO298 (SEQ ID NO:515).

The entire nucleotide sequence of UNQ261 (DNA39975-1210) is shown in Figure 218 (SEQ ID NO:514). Clone DNA39975-1210 contains a single open reading frame with an apparent translational initiation site at nucleotide positions 375-377. The predicted polypeptide precursor is 364 amino acids long. The protein contains four putative transmembrane domains between amino acid positions 36-55 (type II TM), 65-84, 188-208, and 229-245, respectively. A putative N-linked glycosylation site starts at amino acid position 253. In addition, the following features have been identified in the protein sequence: cAMP- and cGMP-dependent protein kinase phosphorylation site, starting at position 8; N-myristoylation sites starting at position 173 and 262, respectively; and a ZP domain between amino acid positions 45-60. Clone DNA39975-1210 has been deposited with ATCC (April 21, 1998) and is assigned ATCC deposit no.209783.

#### EXAMPLE 96: Isolation of cDNA Clones Encoding Human PRO337

A cDNA sequence identified in the amylase screen described in Example 2 above is herein designated DNA42301 (Figure 223, SEQ ID NO:524). The DNA42301 sequence was then compared to other EST

sequences using phrap as described in Example 1 above and a consensus sequence designated herein as DNA28761 was identified. Based on this consensus sequence, oligonucleotides were synthesized: 1) to identify by PCR a cDNA library that contained the sequence of interest, and 2) for use as probes to isolate a clone of the full-length coding sequence. In order to screen several libraries for a source of a full-length clone, DNA from the libraries was screened by PCR amplification with the PCR primer pair identified above. A positive library was then used to isolate clones encoding the PRO337 gene using the probe oligonucleotide and one of the PCR primers. RNA for construction of the cDNA libraries was isolated from human fetal brain.

A cDNA clone was sequenced in its entirety. The full length nucleotide sequence of DNA43316-1237 is shown in Figure 221 (SEQ ID NO:522). Clone DNA43316-1237 contains a single open reading frame with an apparent translational initiation site at nucleotide positions 134-136 (Figure 221; SEQ ID NO:522). The predicted polypeptide precursor is 344 amino acids long. Clone DNA43316-1237 has been deposited with ATCC and is assigned ATCC deposit no. 209487

Based on a BLAST-2 and FastA sequence alignment analysis of the full-length sequence, PRO337 shows amino acid sequence identity to rat neurotrimin (97%).

#### **EXAMPLE 97: Isolation of cDNA Clones Encoding Human PRO403**

##### **Introduction:**

Human thrombopoietin (THPO) is a glycosylated hormone of 352 amino acids consisting of two domains. The N-terminal domain, sharing 50% similarity to erythropoietin, is responsible for the biological activity. The C-terminal region is required for secretion. The gene for thrombopoietin (THPO) maps to human chromosome 3q27-q28 where the six exons of this gene span 7 kilobase base pairs of genomic DNA (Chang et al., Genomics 26: 636-7 (1995); Foster et al., Proc. Natl. Acad. Sci. USA 91: 13023-7 (1994); Gurney et al., Blood 85: 981-988 (1995). In order to determine whether there were any genes encoding THPO homologues located in close proximity to THPO, genomic DNA fragments from this region were identified and sequenced. Three P1 clones and one PAC clones (Genome Systems Inc., St. Louis, MO; cat. Nos. P1-2535 and PAC-6539) encompassing the THPO locus were isolated and a 140 kb region was sequenced using the ordered shotgun strategy (Chen et al., Genomics 17: 651-656 (1993)), coupled with a PCR-based gap filling approach. Analysis reveals that the region is gene-rich with four additional genes located very close to THPO: tumor necrosis factor-receptor type 1 associated protein 2 (TRAP2) and elongation initiation factor gamma (eIF4), chloride channel 2 (CLCN2) and RNA polymerase II subunit hRPB17. While no THPO homolog was found in the region, four novel genes have been predicted by computer-assisted gene detection (GRAIL)(Xu et al., Gen. Engin. 16: 241-253 (1994), the presence of CpG islands (Cross, S. and Bird, A., Curr. Opin. Genet. & Devel. 5: 109-314 (1995), and homology to known genes (as detected by WU-BLAST2.0)(Altschul and Gish, Methods Enzymol. 266: 460-480 (1996) (<http://blast.wustl.edu/blast/README.html>).

##### **Procedures:**

##### **P1 and PAC clones:**

The initial human P1 clone was isolated from a genomic P1 library (Genome Systems Inc., St. Louis, MO; cat. no.: P1-2535) screened with PCR primers designed from the THPO genomic sequence (A.L. Gurney, et al., Blood 85: 981-88 (1995). PCR primers were designed from the end sequences derived from this P1 clone

were then used to screen P1 and PAC libraries (Genome Systems, Cat. Nos.: P1-2535 & PAC-6539) to identify overlapping clones (PAC1, p1.t, and P1.u). The 3'-end sequence from PAC.z was used to define the primers used for the screening of a human BAC library (Genome Systems Inc., St. Louis, MO; Cat. No.: BDTW-4533A).

#### Ordered Shotgun Strategy:

The Ordered Shotgun Strategy (OSS) (Chen et al., Genomics 17: 651-656 (1993)) involves the mapping and sequencing of large genomic DNA clones with a hierarchical approach. The P1 or PAC clone was sonicated and the fragments subcloned into lambda vector ( $\lambda$ Bluestar) (Novagen, Inc., Madison, WI; cat. no. 69242-3). The lambda subclone inserts were isolated by long-range PCR (Barnes, W. Proc. Natl. Acad. Sci. USA 91: 2216-2220 (1994) and the ends sequenced. The lambda-end sequences were overlapped to create a partial map of the original clone. Those lambda clones with overlapping end-sequences were identified, the insets subcloned into a plasmid vector (pUC18 or pUC19, Hoefer Pharmacia Biotech, Inc., San Francisco, CA, Cat. Nos. 27-4949-01 and 27-4951-01) and the ends of the plasmid subclones were sequenced and assembled to generate a contiguous sequence. This directed sequencing strategy minimizes the redundancy required while allowing one to scan for and concentrate on interesting regions.

In order to define better the THPO locus and to search for other genes related to the hematopoietin family, five genomic clones were isolated from this region by PCR screening of human P1 and PAC libraries (Genome System, Inc., Cat. Nos.: P1-2535 and PAC-6539).

The sizes of the genomic fragments are as follows: P1.t is 40 kb; P1.g is 70 kb; P1.u is 70 kb; PAC.z is 200 kb; and BAC.1 is 80 kb. Approximately 75% (140 kb) of the 190 kb genomic DNA region was sequenced by the Ordered Shotgun Strategy (OSS) (Chen et al., Genomics 17: 651-56 (1993), and assembled into contigs using AutoAssembler<sup>TM</sup> (Applied Biosystems, Perkin Elmer, Foster City, CA, cat. no. 903227). The preliminary order of these contigs was determined by manual analysis. There were 47 contigs the 140 kb region. A PCR-based approach to ordering the contigs and filling in the gaps was employed. The following summarizes the number and sizes of the gaps. The 50 kb of sequence unique to BAC.1 was sequenced by a total shotgun approach with a ten-fold redundancy.

<u>Size of gap</u>	<u>number</u>
< 50 bp	13
50-150 bp	7
150-300 bp	7
300-1000 bp	10
1000-5000 bp	7
> 5000 bp	2 ((15,000 bp)

#### DNA sequencing:

ABI DYE-primer<sup>TM</sup> chemistry (PE Applied Biosystems, Foster City, CA; Cat. No.: 402112) was used to end-sequence the lambda and plasmid subclones. ABI DYE-terminator<sup>TM</sup> chemistry (PE Applied Biosystems, Foster City, CA, Cat. No: 403044) was used to sequence the PCR products with their respective PCR primers. The sequences were collected with an ABI377 instrument. For PCR products larger than 1kb, walking primers

were used. The sequences of contigs generated by the OSS strategy in AutoAssemblerTM (PE Applied Biosystems, Foster City, CA; Cat. No: 903227) and the gap-filling sequencing trace files were imported into SequencherTM (Gene Codes Corp., Ann Arbor, MI) for overlapping and editing. The sequences generated by the total shotgun strategy were assembled using Phred and Phrap and edited using Consed (<http://chimera.biotech.washington.edu/uwgc/projects.htm>) and GFP (Genome Reconstruction Manager for Phrap), version 1.2 (<http://stork.cellb.bcm.tmc.edu/gfp/>).

#### PCR-Based gap filling Strategy:

Primers were designed based on the 5'- and 3'-end sequenced of each contig, avoiding repetitive and low quality sequence regions. All primers were designed to be 19-24-mers with 50-70% G/C content. Oligos were synthesized and gel-purified by standard methods.

Since the orientation and order of the contigs were unknown, permutations of the primers were used in the amplification reactions. Two PCR kits were used: first, XL PCR kit (Perkin Elmer, Norwalk, CT; Cat. No.: N8080205), with extension times of approximately 10 minutes; and second, the Taq polymerase PCR kit (Qiagen Inc., Valencia, CA; Cat. No.: 201223) was used under high stringency conditions if smeared or multiple products were observed with the XL PCR kit. The main PCR product from each successful reaction was extracted from a 0.9% low melting agarose gel and purified with the GeneClean DNA Purification kit prior to sequencing.

#### Analysis:

The identification and characterization of coding regions was carried out as follows: First, repetitive sequences were masked using RepeatMasker (A.F.A. Smit & P. Green, [http://ftp.genome.washington.edu/RM/RM\\_details.html](http://ftp.genome.washington.edu/RM/RM_details.html)) which screens DNA sequences in FastA format against a library of repetitive elements and returns a masked query sequence. Repeats not masked were identified by comparing the sequence to the GenBank database using WUBLAST2.0 [Altschul, S & Gish, W., Methods Enzymol. 266: 460-480 (1996); <http://blast.wustl.edu/blast/README.html>] and were masked manually.

Next, known genes were revealed by comparing the genomic regions against Genentech's protein database using the WUBLAST2.0 algorithm and then annotated by aligning the genomic and cDNA sequences for each gene, respectively, using a Needleman-Wunch (Needleman and Wunsch, J. Mol. Biol. 48: 443-453 (1970) algorithm to find regions of local identity between sequences. The strategy results in detection of all exons of the five known genes in the region, THPO, TRAP2, eIF4g, CLCN2 and hRPB17 (see below).

#### Known genes

eukaryotic translation initiation factor 4 gamma  
thrombopoietin  
chloride channel 2  
TNF receptor associated protein 2  
RNA polymerase II subunit hRPB17

#### Map position

3q27-qter  
3q26-q27  
3q26-qter  
not previously mapped  
not previously mapped

Finally, novel transcription units were predicted using a number of approaches. CpG islands (S. Cross & Bird, A., Curr. Opin. Genet. Dev. 5: 109-314 (1995) islands were used to define promoter regions and were

identified as clusters of sites cleaved by enzymes recognizing GC-rich, 6 or 8-mer palindromic sequences (NotI, NarI, BssHII, XhoI. CpG islands are usually associated with promoter regions of genes. WUBLAST2.0 analysis of short genomic regions (10-20 kb) versus GenBank revealed matches to ESTs. The individual EST sequences (or where possible, their sequence chromatogram files) were retrieved and assembled with Sequencer to provide a theoretical cDNA sequence (DNA36443). GRAIL2 (ApoCom Inc., Knoxville, TN, command line version for the DEC alpha) was used to predict a novel exon. The five known genes in the region served as internal controls for the success of the GRAIL algorithm.

#### Isolation:

A partial endothelin converting enzyme-2 (ECE-2) cDNA clone was isolated by first splicing in silico the ECE-2 exons predicted in the genomic sequence to generate a putative sequence (DNA36443). An oligonucleotide probe: GAAGCAGTGCAGCCAGCAGTAGAGAGGCACCTGCTAAGA) (SEQ ID NO:530) was designed and used to screen a human fetal small intestine library (LIB110) and internal PCR primers (36443f1) (ECE2.f:ACGCAGCTGGAGCTGGTCTTAGCA) (SEQ ID NO:531) and (36443r1) (ECE2.r) (GGTACTGGACCCTAGGGCCACAA) (SEQ ID NO:532) were used to confirm clones hybridizing to the probe prior to sequencing. One positive clone was obtained, however this cDNA (DNA49830) represented a partially spliced transcript containing appropriately spliced exons 1 through 6, followed by intron 6 sequence. The oligo dT primer annealed to a poly-A-stretch within an Alu element present in intron 6. A cDNA fragment (DNA49831) was obtained by PCR from a human fetal kidney library (LIB227) with primers designed from the presumed cDNA sequence [36443f3: CCTCCAGCCGAGACCAGTGG (SEQ ID NO:533) and 36443r2: GGTCTTATAAGGGCCAAGACC (SEQ ID NO:534)]. This PCR product extended from exon 13 into the 3' untranslated region in exon 18.

A full length endothelin converting enzyme 2 (ECE-2) cDNA clone (DNA55800-1263) was isolated from an oligo-dT-primed human fetal brain library. RNA from human fetal brain tissue (20 weeks gestation, #283005)(SRC175) was isolated by guanidine thiocyanate and 5 µg used to generate double stranded cDNA which was cloned into the vector pRK5E. The 3' -primer (pGACTAGTTCTAGATCGCGAGCGGCCCTTTTTTTTTTTTTT) (SEQ ID NO:535) and the 5 -linker (pCGGACGCGTGGGTCGA) (SEQ ID NO:536) were designed to introduce XhoI and NotI restriction sites. The library was screened with PCR primers [36443pcrf1: CGGCCGTGATGGCTGGTGACG (SEQ ID NO:537) and 36443r3: GGCAGACTCCTTCCTATGGG (SEQ ID NO:538)] designed from the partial human ECE-2 cDNA sequences (DNA49830 and DNA49831). PCR products were cloned into the vector pCR2.1-TOPO (Invitrogen Corp., Carlsbad, CA, Cat. No. K4500-01) and sequenced with DYE-terminator chemistry as described above.

#### EXAMPLE 98: Northern Blot and in situ RNA Hybridization Analysis for PRO403

Expression of PRO403 mRNA in human tissues was examined by Northern blot analysis. Human polyA+ RNA blots derived from human fetal and adult tissues (Clontech, Palo Alto, CA; Cat. Nos. 7760-1, 7756-1 and 7755-1) were hybridized to a [32P-α]dATP-labelled cDNA fragments from probe based on the full length PRO403 cDNA. Blots were incubated with the probes in hybridization buffer (5X SSPE; 2X Denhardt's solution; 100 mg/mL denatured sheared salmon sperm DNA; 50% formamide; 2% SDS) for 18 hours at 42°C,

washed to high stringency (0.1XSSC, 0.1% SDS, 50°C) and autoradiographed. The blots were developed after overnight exposure by phosphorimager analysis (Fuji).

PRO403 mRNA transcripts were detected. Analysis of the expression pattern showed the strongest signal of the expected 3.3 kb transcript in adult brain (highest in the cerebellum, putamen, medulla, and temporal lobe, and lower in the cerebral cortex, occipital lobe and frontal lobe), spinal cord, lung and pancreas and higher levels of a 4.5 kb transcript in fetal brain and kidney.

#### EXAMPLE 99: Use of PRO Polypeptide-Encoding Nucleic Acid as Hybridization Probes

The following method describes use of a nucleotide sequence encoding a PRO polypeptide as a hybridization probe.

DNA comprising the coding sequence of a PRO polypeptide of interest as disclosed herein may be employed as a probe or used as a basis from which to prepare probes to screen for homologous DNAs (such as those encoding naturally-occurring variants of the PRO polypeptide) in human tissue cDNA libraries or human tissue genomic libraries.

Hybridization and washing of filters containing either library DNAs is performed under the following high stringency conditions. Hybridization of radiolabeled PRO polypeptide-encoding nucleic acid-derived probe to the filters is performed in a solution of 50% formamide, 5x SSC, 0.1% SDS, 0.1% sodium pyrophosphate, 50 mM sodium phosphate, pH 6.8, 2x Denhardt's solution, and 10% dextran sulfate at 42°C for 20 hours. Washing of the filters is performed in an aqueous solution of 0.1x SSC and 0.1% SDS at 42°C.

DNAs having a desired sequence identity with the DNA encoding full-length native sequence PRO polypeptide can then be identified using standard techniques known in the art.

#### EXAMPLE 100: Expression of PRO Polypeptides in *E. coli*

This example illustrates preparation of an unglycosylated form of a desired PRO polypeptide by recombinant expression in *E. coli*.

The DNA sequence encoding the desired PRO polypeptide is initially amplified using selected PCR primers. The primers should contain restriction enzyme sites which correspond to the restriction enzyme sites on the selected expression vector. A variety of expression vectors may be employed. An example of a suitable vector is pBR322 (derived from *E. coli*; see Bolivar et al., *Gene*, 2:95 (1977)) which contains genes for ampicillin and tetracycline resistance. The vector is digested with restriction enzyme and dephosphorylated. The PCR amplified sequences are then ligated into the vector. The vector will preferably include sequences which encode for an antibiotic resistance gene, a trp promoter, a polyhis leader (including the first six STII codons, polyhis sequence, and enterokinase cleavage site), the specific PRO polypeptide coding region, lambda transcriptional terminator, and an argU gene.

The ligation mixture is then used to transform a selected *E. coli* strain using the methods described in Sambrook et al., *supra*. Transformants are identified by their ability to grow on LB plates and antibiotic resistant colonies are then selected. Plasmid DNA can be isolated and confirmed by restriction analysis and DNA sequencing.

Selected clones can be grown overnight in liquid culture medium such as LB broth supplemented with antibiotics. The overnight culture may subsequently be used to inoculate a larger scale culture. The cells are then grown to a desired optical density, during which the expression promoter is turned on.

After culturing the cells for several more hours, the cells can be harvested by centrifugation. The cell pellet obtained by the centrifugation can be solubilized using various agents known in the art, and the solubilized

PRO polypeptide can then be purified using a metal chelating column under conditions that allow tight binding of the protein.

PRO181, PRO195, PRO200, PRO237, PRO273, PRO540, PRO322, PRO1017, PRO938, PRO162, PRO1114, PRO827 and PRO1008 were expressed in *E. coli* in a poly-His tagged form, using the following procedure. The DNA encoding the PRO polypeptide was initially amplified using selected PCR primers. The primers contained restriction enzyme sites which correspond to the restriction enzyme sites on the selected expression vector, and other useful sequences providing for efficient and reliable translation initiation, rapid purification on a metal chelation column, and proteolytic removal with enterokinase. The PCR-amplified, poly-His tagged sequences were then ligated into an expression vector, which was used to transform an *E. coli* host based on strain 52 (W3110 fuhA(tonA) lon galE rpoHts(htpRts) clpP(lacIq). Transformants were first grown in LB containing 50 mg/ml carbenicillin at 30°C with shaking until an O.D.600 of 3-5 was reached. Cultures were then diluted 50-100 fold into CRAP media (prepared by mixing 3.57 g (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 0.71 g sodium citrate-2H<sub>2</sub>O, 1.07 g KCl, 5.36 g Difco yeast extract, 5.36 g Sheffield hycase SF in 500 mL water, as well as 110 mM MPOS, pH 7.3, 0.55% (w/v) glucose and 7 mM MgSO<sub>4</sub>) and grown for approximately 20-30 hours at 30°C with shaking. Samples were removed to verify expression by SDS-PAGE analysis, and the bulk culture is centrifuged to pellet the cells. Cell pellets were frozen until purification and refolding.

*E. coli* paste from 0.5 to 1 L fermentations (6-10 g pellets) was resuspended in 10 volumes (w/v) in 7 M guanidine, 20 mM Tris, pH 8 buffer. Solid sodium sulfite and sodium tetrathionate is added to make final concentrations of 0.1M and 0.02 M, respectively, and the solution was stirred overnight at 4°C. This step results in a denatured protein with all cysteine residues blocked by sulfitolization. The solution was centrifuged at 40,000 rpm in a Beckman Ultracentrifuge for 30 min. The supernatant was diluted with 3-5 volumes of metal chelate column buffer (6 M guanidine, 20 mM Tris, pH 7.4) and filtered through 0.22 micron filters to clarify. Depending the clarified extract was loaded onto a 5 ml Qiagen Ni-NTA metal chelate column equilibrated in the metal chelate column buffer. The column was washed with additional buffer containing 50 mM imidazole (Calbiochem, Utrrol grade), pH 7.4. The protein was eluted with buffer containing 250 mM imidazole. Fractions containing the desired protein were pooled and stored at 4°C. Protein concentration was estimated by its absorbance at 280 nm using the calculated extinction coefficient based on its amino acid sequence.

The proteins were refolded by diluting sample slowly into freshly prepared refolding buffer consisting of: 20 mM Tris, pH 8.6, 0.3 M NaCl, 2.5 M urea, 5 mM cysteine, 20 mM glycine and 1 mM EDTA. Refolding volumes were chosen so that the final protein concentration was between 50 to 100 micrograms/ml. The refolding solution was stirred gently at 4°C for 12-36 hours. The refolding reaction was quenched by the addition of TFA to a final concentration of 0.4% (pH of approximately 3). Before further purification of the protein, the solution was filtered through a 0.22 micron filter and acetonitrile was added to 2-10% final concentration. The refolded protein was chromatographed on a Poros R1/H reversed phase column using a

## FIGURE 89

></usr/seqdb2/sst/DNA/Dnaseqs.min/ss.DNA50919  
><subunit 1 of 1, 472 aa, 1 stop  
><MW: 53847, pI: 5.75, NX(S/T): 2  
MSNIYIQEPPTNGKVLLKTTAGDIDIELWSKEAPKACRNFIQLCLEAYDNTIFHRVVPGFI  
VQGGDPTGTGSGGESIYGAPFKDEFHSRLRFNRRGLVAMANAGSHDNGSQFFFTLGRADELN  
NKHTIFGKVTGDTVYNMLRLSEVIDDDERPHNPHKIKSCEVLFPNPFDDIIPREIKRLKKEK  
PEEEVKKLKPKGKTNFSLLSPGEEAEVEEVEVNRVSQSMKGKSSSHDLLKDDPHLSSVPVV  
ESEKGDAPDLVDDGEDESAEHDEYIDGDEKNLMRERIAKKLKKDTSANVKSAGEGEVEKKSV  
SRSEELRKEARQLKRELLAAKQKKVENAAKQAEKRSEEEAAPPDGAVAEYRREKQKYEALRK  
QQSKKGTSREDQTLALLNQFKSKLTQAIAETPENDIPETEVEDDEGWMSHVLQFEDKSRKVK  
DASMQSDSTFEIYDPRNPFVNKRREESKKLMREKKERR

### **Important features:**

#### **Signal peptide:**

amino acids 1-21

#### **N-glycosylation sites.**

amino acids 109-112 and 201-204

#### **Cyclophilin-type peptidyl-prolyl cis-trans isomerase signature.**

amino acids 49-66

#### **Homologous region to Cyclophilin-type peptidyl-prolyl cis-trans isomerase**

amino acids 96-140, 49-89 and 22-51

## FIGURE 90

CGCCGCGCTTGGGGCTGGAAGTTCCTCCGACAGGTCCGTGCCGGGCGAGAGAGATGCTGCCCGG  
CCGCGCTTCGGCTTTGAGGCGAGAGAAGTGTCCACAGCCATTTCGCGCTTGCTGACGGCGCTCG  
AGCCCTGGCCAGACATGTCGCCACAGGGTTCTCCTTCGGGTCCGGGACTCTGGGGCTCCACCACC  
GTGGCCCGCGGCGGGACCAGCACAGGCGGCGTTTCTCCTTCGGAACGGGAAACGTCTAGCAA  
CCCTTCTGTGGGGCTCAATTTTGGAAATCTTGAAGTACTTCACTCCAGCAACTACATCTG  
CTCCTTCAAGTGGTTTGGAAACGGGGCTCTTTGGATCTAAACCTGCCACTGGGTTCACTCTA  
GGAGGAACAAATA CAGGTGCCCTTGCACACCAAGAGGCTCAAGTGGTCAACAAATATGGAAC  
CCTGCAAGGAAAACAGATGCATGTGGGGAAGACACCCATCCAGTCTTTTAGGAGTCCCTT  
TCTCCAGACCTCCTCTAGTATCTCAGGTTTGCACCTCCAGAACCCCGAGCCCTGGAAA  
GGAATCAGAGATGCTACCACTACCGCCTGGATGGAGTCTCGCTCTGTGCCAGGCTGGAG  
TGCAGTGGCAGATCTCGGCTCACTGCAACCTCCGCTCCCGGGTTCAAGCGAGTCTCCTGC  
CTCAGCCTCTGAGTGTCTGGGGCTACAGGTGCCTGCAGGAGTCTCTGGGGCCAGCTGGGCTCG  
ATGTACGTTCAGCACCGCGGGAACGGTACAAGTGGCTGCGCTTCAGCGAGGACTGTCTGTACT  
GAACGTGTACCGCGCGGCGCGCGCCGGGATCCCAAGTGCAGTGATGGTCTGGTTCC  
CGGGAGGCGCCTTCATCTGTGGGCGCTGCTTCTTCGTACGAGGGCTCTGACTTGGCCGCGCCG  
GAGAAAGTGGTGTGGTGTTCCTGCAGCACAGGCTCGGCATCTTCGGCTTCTGAGCACGGA  
CGACAGCCACGCGCGCGGGAACCTGGGGCTGCTGGACCAGATGGCGCTCTGCGCTGGGTGC  
AGGAGAACATCGCAGCCTTCGGGGGAGACCCAGGAAATGTGACCCCTGTTCCGCCAGTCCGGC  
GGGGCCATGAGCATCTCAGGACTGATGATGTCACCCCTAGCCTCGGGTCTCTTCCATCGGGC  
CATTTCCAGAGTGGCACCGCGTTATTAGACTTTTCATCACTAGTAACCCACTGAAAGTGG  
CCAAGAAGGTTGCCACCTGGCTGGATGCAACCACAAAGCACACAGATCCTGGTAAACTGC  
CTGAGGGCACTATCAGGGACCAAGGTGATGCGTGTGTCCAAAGATGAGATTCTTCCAAC  
GAACCTCCAGAGAGACCCGGAAGAGATTATCTGGTCCATGAGCCCTGTGGTGGATGGTGTGG  
TGATCCAGAGTACCTTTGGTGTCTCTGACCCAGGGGAAGGTTTCATCTGTGCCCTACCTT  
CTAGGTGTCAACAACCTGGAATTCAATTGGCTCTTGCCCTTATAATATCACCAGGAGCAGGT  
ACCACTTGTGGTGGAGGAGTACCTGGACAATGTCAATGAGCATGACTGGAAGATGCTACGAA  
ACCGTATGATGGACATAGTTCAAGATGCCACTTTCGTGTATGCCACACTGCAGACTGCTCAC  
TACCAACCGAGAAACCCCAATGATGGGAATCTGCCCTGCTGGCCACGCTACAACAAGGATGAA  
AAGTACCTGCAGCTGGATTTTACCAAGAGTGGGCATGAGCTCAAGGAGAAGAAGATGGC  
TTTTTGGATGAGTCTGTACCAGTCTCAAAGACCTGAGAAGCAGAGGCAATTCTAAGGGTGGC  
TATGCAGGAAGGAGCCAAAGAGGGGTTTGCCCCCACCATCCAGGCCCTGGGGAGACTAGCCA  
TGGACATACCTTGGGGACAAGAGTTCTACCCACCCAGTTTGAAGCTGCAGGAGCTCCCTGCT  
GCCTCCAGGCCAAAGCTAGAGCTTTTGCTGTGTGTGGGACCTGCCTGCGCTTTCAGGCC  
TGACATCCCATGATGCCCTCTACTTCACTGTTGACATCCAGTTAGGCCAGGCCCTGTCAAC  
ACCACATGTGCTCAGCTCTCCAGCCTCAGGACAACTCTTTTTTCCCTTCTTCAAATCCT  
CCCACCTTCAATGTCTCCTTGTGACTCCTTCTTATGGGAGGTGCAGCCAGACTGCCACTGC  
CCCTGTCACTGCACCCAGCTTGGCATTACCATCCATCCCTGCTCAACCTTGTTCCTGTCTGT  
TCACATTGGCCTGGAGCCTTAGGGCAGGTGTGACATGGAGCAAACTTTTGGTAGTTTGGGA  
TCTTCTCTCCACCCACACTTATCTCCCCAGGGCCACTCAAAGTCTATACACAGGGGTG  
TCTCTTCAATAAAGAGTGTGATTAGAAAAA

## FIGURE 91

</usr/seqdb2/sst/DNA/Dnaseqs.min/ss.DNA44179  
<subunit 1 of 1, 545 aa, 1 stop  
<MW: 58934, pI: 9.45, NX(S/T): 4  
MSTGFSFGSGTLGSTTVAAGGTSTGGVFSFGTGTSSNPSVGLNFGNLGSTSTPATTSA PSSG  
FGTGLFGSKPATGFTLGGTNTGALHTKR PQVVTKYGT LQ GKQMHVGTPIQVFLGV PFSRPP  
LGILRFAPPEPEPWKGIRDATTYPGWSLALSPGWSAVARSRLTATSASRVQASLLPQPLS  
VWGYRCLQESWGQLASMYVSTRERYKWLRFSEDCLYLNVYAPARAPGDPQLPVMVWFPGGAF  
IVGAASSYEGSDLAAREKVVLVFLQHR LGIFGFLSTDDSHARGNWGLLDQMAALRWVQENIA  
AFGGDPGNVT LFGQSAGAMSI SGLMMSPLASGLFHRAISQSGTALFRLFITSNPLKVAKKVA  
HLAGCNHNSTQILVNCLRALSGTKVMRVSNKMRFLQLNFQRPDEEIIWSMSPVVDGVVIPDD  
PLVLLTQGKVSSVPYLLGVNNLEFNWLLPYNITKEQVPLVVEEYLDNVNEHDWKMLRNRMMD  
IVQDATFVYATLQTAHYHRET PMMGICPAGHATTRMKSTCSWILPQEWA

### **Important features:**

#### **Signal peptide:**

amino acids 1-29

#### **Carboxylesterases type-B serine active site.**

amino acids 312-327

#### **Carboxylesterases type-B signature 2.**

amino acids 218-228

#### **N-glycosylation sites.**

amino acids 318-321, 380-383 and 465-468

# FIGURE 92

GAGAAACAGGCTGTCTCAGGCAGGCCCTGCGCCTCCTATGCGGAGATGCTACTGCCACTGCT  
 GCTGTCTCTCGTGTGGGCGGGTCCAGGCTATGGATGGGAGATTCTGGATACGAGTGCAGG  
 AGTCAGTGATGGTGCCGGAGGGCTGTGCATCTCTGTGCCCTGCTCTTTCTCCATCCCCGA  
 CAAGACTGGACAGGGTCTACCCAGCTTATGGCTACTGGTTCAAAGCAGTGACTGAGACAAC  
 CAAGGGTGTCTCTGTGGCCACAAACCACAGAGTCGAGAGGTGGAATGAGCACCCGGGGCC  
 GATTCAGCTCACTGGGGATCCGCCAAGGGAACTGCTCCTTGGTGATCAGAGACGCGCAG  
 ATGCAGGATGAGTCACTATCTCTTTCGGGTGGAGAGGAAGCTATGTGACATATAATTT  
 CATGAACGATGGGTTCTTTCTAAAAGTAACAGTGCTCAGCTTCAGCCCCAGACCCAGGACC  
 ACAACACCGACTCACTGCGCATGTGGACTTCTCCAGAAAGGGTGTGAGCGCACAGAGGACC  
 GTCGACTCCGTGTGGCCCTATGCCCCCAGAGACCTTGTTATCAGCATTTACAGTGACAACAC  
 GCCAGCCTTGGAGCCCCAGCCCCAGGGAATGTCCCATACCTTGAAGGCCCAAAAGGCCAGT  
 TCCTGCGGCTCCTCTGTGCTGTGACAGCCAGCCCCCTGCCACACTGAGCTGGGTCTGTGAG  
 AACAGAGTCTCTCTCTGCTCCATCCCTGGGGCCCTAGACCCCTGGGGCTGGAGCTGCCCGG  
 GGTGAAGGCTGGGATTCAGGGCGCTACACCTGCCGAGCGGAGAACAGGCTTGGCTCCAGC  
 AGCGAGCCTTGGACCTCTCTGTGCTGATCTCTCCAGAGAACCTGAGAGTGATGGTTTCCCAA  
 GCAACAGGACAGTCTCTGGAACACCTTGGGAACGGCAGCTCTCTCCAGTACTGGAGGCCA  
 AAGCCTGTGCCCTGGTCTGTGTACACACAGCAGCCCCCAGCCAGCTGAGCTGGACCGCAGA  
 GGGGACAGGTTCTGAGCCCCCTCCAGCCCTCAGACCCCGGGGTCTTGGAGCTGCTCTGGGT  
 CAAGTGGAGCAGGAAGGAGAGTTACCTGCCAGCTCGGCACCCACTGGGCTCCAGCAGCT  
 CTCTCTCAGCTCTCCGTGCACTATAAGAAGGGACTCATCTCAACGGGATTCTCCAACGGAG  
 CGTTTCTGGGAATCGGCATCAGGCTCTCTTTTCTCTGCTGGCCTGTATCATCATGAAG  
 ATTCTACCGAAGAGACGGACTCAGACAGAAACCCGAGGCCAGGTTCTCCGGCACAGCAC  
 GATCCTGGATTACATCAATGTGGTCCCGACGGCTGGCCCCCTGGCTCAGAAGCGGAATCAGA  
 AAGCCACCAAAACAGTCTCGGACCCCTCTCCACAGGTGCTCCTCTCCCAAGCAATCAAG  
 AAGAACCAGAAAAAGCAGTATCAGTTGCCAGTTTCCCAAGAACCAATCATCCACTCAAGC  
 CCCAGAATCCAGGAGAGCCAAAGAGGAGCTCCATTATGCCACGCTCAACTTCCAGGCGTCA  
 GACCCAGGCTTGGGCCCCGATGCCCAAGGGCACCCAGCGGATTATGCAGAAGTCAAGTTT  
 CAATGAGGGTCTCTTAGGCTTTAGGACTGGGACTTCGGCTAGGGAGGAAGGTAGAGTAAGAG  
 GTTGAAGATAACAGAGTGCAAAAGTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT  
 CTCTCTCTCTCTCTCTTTAAAAAAACATCTGGCCAGGGCACAGTGGCTCACGCCCTGTAATC  
 CCAGCACTTTGGGAGGTTGAGGTGGGCAGATCGCCTGAGTCTGGGAGTTTCAGACCCAGCCTG  
 GCCAACTTGGTGAACCCCGTCTCTACTAAAAATACAAAAATTAGCTGGGCATGGTGGCAGG  
 CGCTGTAACTCTACTTGGGAAGCTGAGGCAGGAGAATCACTTGAACCTGGGAGACGG  
 AGGTTGCAGTGAGCCAAGATCACACCATTGACAGCCAGCTGGGCAACAAAGCGAGACTCCA  
 TCTCAAAAAAAAATCTCCAAATGGGTTGGGTGTCTGTAATCCAGCACTTTGGGAGGCTA  
 AGTGGGTGGATTGCTTGAAGCCAGGAGTTTCAGACCCAGCTGGGCAACATGGTGAACCCCT  
 ATCTCTACAAAAAATACAAAAATAGCTGGGCTTGGTGGTGTGTCCTGTAGTCCAGCTGT  
 CAGACATTTAAACCAGAGCAACTCCATCTGGAATAGGAGCTGAATAAAATGAGGCTGAGACC  
 TACTGGGCTGCACTCTCAGACAGTGGAGGCACTTAAGTCAAGAGTGAAGACAGGAGGTCG  
 TACAAGATACAGGCTATAAAGACTTTTGCTGATAAAAAAGATTCAGTGAAGAAAGCAACAA  
 ATCCCAACCAAAACCAAGTTGGCCACGAGAGTGACCTCTGGTCTGCTCACTGCTACACTCT  
 GACAGCACCATGACAGTTTACAAATGCCATGGCAACATCAGGAAGTACCAGATATGTCCCA  
 AAAGGGGGAGGAATGAATAATCCACCCCTTGTGTAGCAATAAGCAAGAAATAACCATAAAA  
 GTGGGCAACACAGAGCTCTAGGCGTGCTCTGTCTATGGAGTAGCCATCTTTTGTCTCT  
 TACTTTCTTAATAAACTTGCTTTACCTTAAAAAAA

10017001-10017001

## **FIGURE 93**

></usr/seqdb2/sst/DNA/Dnaseqs.min/ss.DNA54002  
><subunit 1 of 1, 544 aa, 1 stop  
><MW: 60268, pI: 9.53, NX(S/T): 3  
MLLPLLSSLLGGSQAMDGRFWIRVQESVMVPEGLCISVPCSFYSYPRQDWTGSTPAYGYWFK  
AVTETTKGAPVATNHQSREVEVMSTRGRFQLTGDDPAKGNC SLVIRDAQMQDESQYFFRVERGS  
YVTYNFMNDGFFLKVTVLSTPRPQDHNTDLTCHVDFSRKGVSAQRTVRLRVAYAPRDLVIS  
ISRDNTPALEFPQPQGNVPYLEAQKGQFLRLCAADSQPPATLSWVLQNRVLSSSHPWGPRPL  
GLELPGVKAGDSGRYTCRAENRLGSQQRALDLSVQYPENLRVMVSQANRTVLENLNGTSL  
PVLEGQSLCLVCVTHSSPPARLSWTQRGQVLSPSQSPDPGVLELPRVQVEHEGEFTCHARHP  
LGSQHVSLSLSVHYHKGLISTAFSNGAFLGIGITALLFLCLALIIMKILPKRRTQTETPRPR  
FSRHSTILDYINVVPTAGPLAQKRNQKATPNSPRTPPPPGAPSPESKKNQKKQYQLPSFPEP  
KSSTQAPESQESQEELHYATLNFPGVRPRPEARMPKGTQADYAEVKFQ

### **Important features:**

#### **Signal peptide:**

amino acids 1-15

#### **Transmembrane domain:**

amino acids 399-418

#### **N-glycosylation site.**

amino acids 100-103, 297-300 and 306-309

#### **Immunoglobulins and major histocompatibility complex proteins signature.**

amino acids 365-371

# FIGURE 94

TGAAGAGTAATAGTTGGAATCAAAGAGTCAACGCAATGAACTGTTATTTACTGCTGCGTTT  
TATGTTGGGAATTCTCTCTCTATGGCCTTGCTCTGGAGCAACAGAAAACCTCTCAACAAAGA  
AAGTCAAGCAGCCAGTGCATCTCATTGAGAGTGAAGCGTGCTGGGTGGAACCAATTT  
TTTGTACCAGAGGAAATGAATACGACTAGTCATCACATCGGCCAGCTAAGATCTGATTAGA  
CAATGGAAACAATTCTTCCAGTACAAGCTTTTGGGAGCTGGAGCTGGAAGTACTTTTATCA  
TTGATGAAGAACAAGTGCATATATGCCATAAGCTTGATAGAGAGGAGCGGATCCCTC  
TACATCTTAAGAGCCAGGTAATAGACATCGCTACTGGAAGGGCTGTGGAAACCTGAGTCTGA  
GTTTGTCTCAAAAGTTTCCGATATCAATGACAATGAACCAAAATCTCTAGATGAACCTTATG  
AGGCCATTGTACCAGAGATGTCTCCAGAAGGAACATTAGTTATCCAGGTGACAGCAAGTGAT  
GCTGACGATCCCTCAAGTGGTAATAATGCTCGTCTCTCTACAGCTTACTCTCAAGGCCAGCC  
ATATTTTTCTGTTGAACCAACAACAGGAGTCATAAGAATATCTTCTAAATGGATAGAGAAC  
TGCAAGATGAGTATTGGGTAATCATTCAAGCCAAGGACATGATTGGTCAGCCAGGAGCGTTG  
TCTGGAACAACAAGTGTTAATTAACCTTTCTAGATGTTAATGACAATAAGCCTATATTTAA  
AGAAAGTTTATACCGCTTGACTGTCTCTGAATCTGCACCCACTGGGACTTCTATAGGAACAA  
TCAATGGCATATGATAATGACATAGGAGAGAATGCAGAAATGGATTAGACGATTGAAGAGGAT  
GATTCGCAAAACATTTGACATTATTAATAATCATGAAACTCAAGAAGGAATAGTTATTTAA  
AAAGAAAGTGGATTTTGAGCACCAGAACCACTACGGTATTAGAGCAAAAGTTAAAAACCATC  
ATGTTCTCGAGCAGCTCATGAAGTACCACACTGAGGCTTCCACCATTCTCATTAAGATCCAG  
GTGGAAGATGTTGATGAGCTCCTCTTTCTCTTCCATATATGTTATTTGAAGTTTGTGA  
AGAAACCCACAGGGATCATTTGTAGGCGTGGTGTCTGCCACAGACCAGACAATAGGAAAT  
CTCCTATCAGGTATTCTATTACTAGGAGCAAGTGTTCAATATCAATGATAATGGTACAATC  
ACTACAAGTAATCTCTGGATCGTGAATCAGTGCTGGTACAACCTTAAGATTACAGCCAC  
AGAAAAATACAATATAGAACAGATCTCTTCGATCCCACTGTATGTGCAAGTTCTTAACATCA  
ATGATCATGCTCCTGAGTTCTCTCAATACTATGAGACTTATGTTTGTGAAAAATGCAGGCTCT  
GGTCAGGTAATTCAGACTATCAGTGCAGTGAGATAGAGATGAATCCATAGAAGAGCACCATT  
TTACTTTAATCTATCTGTAGAAGACATAACAATTCAGGTTTACAATCATAGATAATCAAG  
ATAACACAGCTGTCAATTTTGACTAATAGAACTGGTTTAAACCTCAAGAAGAACCTGTCTTC  
TACATCTCCATCTTAATTCGCCACAATGGAATCCCGTCACTTACAAGTACAACACCCCTTAC  
CATCCATGCTCTGTGACTGTGGTGACAGTGGGAGCACACAGACCTGCCAGTACCAGGAGCTTG  
TGCTTTCCATGGGATTCAAGACAGAAGTTATCATTTGCTATTCTCATTGTCAATTATGATGATA  
TTTGGGTTTATTTTTTGACTTTGGGTTTAAAAACAACGGAGAAAAACAGATTCTATTTCTGTA  
GAAAAGTGAAGATTTTCAAGAGAAATATATTTCCAATATGATGATGAAGGGGTGGAGAAGAAG  
ATACAGAGGCTTTTGATATAGCAGAGCTGAGGAGTAGTACCATAATCGCGGAACGCAAGACT  
CGGAAAAACCAAGCGCTGAGATCAGGAGCCTATACAGGCAGTCTTTGCAAGTTGGCCCCGA  
CAGTGCCATATTCAGGAAATTTTATTTGAGAAAGCTCGAAGAAAGCTAATCTGATCCGTTGTG  
CCCCTCCTTTGATTCCCTCCAGACCTACGCTTTTGAGGGAAACAGGGTCATTAGCTGGATCC  
CTGAGCTCTTGAATCAGCAGCTCTCTGATCAGGATGAAGGCTATGATTACCTTATGAGTT  
GGGACCTCGCTTTAAAGATTAGCATGCGATTTTGGTTCTGCAAGTGAGTCAAAATAATTAGG  
GCTTTTACCATCAAAATTTTAAAGTGCTAATGTGATTGCAACCCAATGGTATGCTTAA  
AGAGTTTGTGCCCTGGCTTTATGCGGGGAAAGCCCTAGTCTATGGAGTTTCTGAGTTTCC  
CTGGAGTAAATACTCCATGGTTATTTTAAAGTACCTACATGCTGTCTATTGAACAGAGATGTG  
GGGAGAAATGTAAACCAATCAGCTCACAGGCTCAATACAACAGATTGAAAGTAAATTAATG  
TAGGAAGATATTAAGTAGATGAGAGGACACAAGATGTAGTCGATCCTTATGCGATTATAT  
CATTTATTACTTAGGAAGAGTAAAAATACCAACGAGAAAAATTTAAAGGAGCAAAATTTG  
CAAGTCAATAGAAATGTAACAAATCAGATAACATTACATTTCTATCATATTGACATGAAA  
ATTGAAAAATGTATAGTCAGAGAAATTTTATGAATTTATTCATGAGTATGTTTCTTTAT  
TTAAA

## **FIGURE 95**

></usr/seqdb2/sst/DNA/Dnaseqs.min/ss.DNA53906  
><subunit 1 of 1, 772 aa, 1 stop  
><MW: 87002, pI: 4.64, NX(S/T): 8  
MNCYLLLRFMLGIPLLWPCLGATENSQTKKVKQPVRSHLRVKRGWVWNQFFVPEEMNTTSHH  
IGQLRSDLNDNGNSFQYKLLGAGAGSTFIIDERTGDIYAIQKLDREERSLYILRAQVIDIAT  
GRAVEPESEFVIKVSNDINDNEPKFLDEPYEAIVPEMSPEGLTLVIQVTASDADDPSGNNARL  
LYSLLQGQPYSVEPTTGVIRISSKMDRELQDEYWVIIQAKDMIGQPGALSGETTSVLIKLSD  
VNDNKPIFKESLYRLTVSESAPTGTSGITIMAYDNDIGENAEMDYSIEEDDSQTFDIITNHE  
TQEGIVILKKKVDFEHQNHYGIRAKVKNHHVPEQLMKYHTEASTTFIKIQVEDVDEPPLFL  
PYVFEVFEETPQGSFVGVVSATDPDNRKSPIRYSITRSKVFVNINDNGTITTSNSLDREISA  
WYNLSITATEKYNIEQISSIPLYVQVLNINDHAPEFSQYYETYVCENAGSGQVIQTIASVDR  
DESIEEHFFYFNLSVEDTNNSSFTIIDNQDNTAVILTNRTGFNLQEEPVFYISILIADNGIP  
SLTSTNTLTIIHVDCDGSSTQTCQYQELVLSMGFKTEVIIAILICIMIIFGFIFLTGLKQ  
RRKQILFPEKSEDFRENIFQYDDEGGGEEDTEAFDIAELRSSTIMRERKTRKTTSAIRSLY  
RQSLQVGPDSAIFRKFILEKLEANTDPCAPPFDSLQTYAFEGTGLAGSLSSLESASVSDQD  
ESYDYLNELGPRFKRLACMFGSAVQSNN

### **Important features:**

#### **Signal peptide:**

amino acids 1-21

#### **Transmembrane domain:**

amino acids 597-617

#### **N-glycosylation sites.**

amino acids 57-60, 74-77, 419-423, 437-440, 508-511, 515-518,  
516-519 and 534-537

#### **Cadherins extracellular repeated domain signature.**

amino acids 136-146 and 244-254



# FIGURE 97

GCAACCTCAGCTTCTAGTATCCAGACTCCAGCGCCGCCCGGGCGCGGACCCCAACCCCGAC  
 CCAGAGCTTCTCCAGCGGCGCGCAGCGAGCAGGGCTCCCCGCCCTTAACCTTCTCCGCGGGG  
 CCCAGCACCTTCCGGAGTCCGGGTTGCCACCTGCAAACTCTCCGCTTCTGACACTTGCCA  
 CCCCAGAGCCAGCGCGGGCCCCGAGCGAGTCTATGGCCAACGCGGGGTGCGAGCTGTTGGGC  
 TTCATCTCGCCTTCTCGGATGGATCGGCGCCATCGTCAGCACTGCCTCGCCCCAGTGGAG  
 GATTTACTCCTATGCCGGCGCAACATCGTGACCGCCAGGCCATGTACGAGGGGCTGTGGA  
 TGTCTCGCTGTGCGCAGAGCACCGGGCAGATCCAGTGCAGAACTCTTTGACTCCTTGCTGAAT  
 CTGACGACCATTTGCAAGCAACCCGTGCTTGTATGGTGTGCTGCTCCTCTGGAGTGAT  
 AGCAATCTTTGTGGCCACCCTTGGCATGAAGTGTATGAAGTGCTTGGGAAGACGATGAGGTGC  
 AGAAGATGAGGATGGCTGTCTATTGGGGGTGCGATATTTCTTCTTGCAAGTCTGGCTATTTTA  
 GTTGCCACAGCATGGTATGGCAATAGAACTCGTTCAAGAATTCTATGACCCCTATGACCCAGT  
 CAATGCCAGGTACGAATTTGGTCAGGCTCTCTTCACTGGCTGGGCTGCTGCTTCTCTGCCC  
 TTCTGGGAGGTGCTTCTTGTCTGTTCTCTGCCCCGAAAAACAACCTCTTACCCAAACCCA  
 AGGCCCTATCCAAAACCTGCACCTTCCAGCGGGAAGACTACGTTGACACAGAGGCAAAAG  
 GAGAAAATCATGTTGAAACAAACCGAAAATGGACATTGAGATACTATCATTAAACATTAGGAC  
 CTTAGAAATTTGGGTATTGTAATCTGAAGTATGTTATCAAAAACAAACAAACAAACAAAA  
 ACCCATGTGTTTAAATACTCAGTGCTAAACATGGCTTAATCTTATTTTATCTTCTTCTCA  
 ATATAGGAGGGGAAATTTTCCATTGTATTACTGCTTCCATTGAGTAATCTACTCAAT  
 GGGGAGGGGTGCTCCTTAAATATATATAGATATGTATATACATCTTCTATTAAAA  
 ATAGACAGTAAATACATATTCTCATTATGTTGATAGTACTAGCATACTTAAATATCTTAAAT  
 AGGTAATGTATTTAATCCATATTGATGAAGATGTTTATTGGTATATTTCTTTTCTGTC  
 TTATATACATATGTAAACGCAAAATATCATTTACTCTTCTTATAGCTTTGGGGTGCCCTTG  
 CCACAAGACCTAGCCTAATTTACCAAGGATGAATTTCTTCAATTTCTCATGCGTGCCCTTT  
 CATATACTTATTTTATTTTACCATAATCTTATAGCACTTGCACTCGTTATTAAGCCCTTAT  
 TTGTTTGTGTTTCTTGGTCTCTATCTCTGAATCTAACACATTTTCATAGCCTACATTTTA  
 GTTTCTAAAGCCAAGAAGATTTTATCAAAATCAGAACTTTGGAGGCAAACTCTTCTGCATG  
 ACCAAAGTGATAAATTTCTGTTGACCTTCCACACAATCCCTGTACTGTGACCCATAGCACT  
 CTTGTTTGTCTTTGAAAATATTTGTCCAATTGAGTAGCTGCATGCTGTTCCCCAGGTTGT  
 AACACACTTTATTTGATTGAATTTTAACTACTTATTATAGTATTTATATCCCCCTAACT  
 ACCTTTTGTGTTCCCATTCCTTAATTGTATTGTTTCCCAAGTGTAATTTATCATGCGTTTA  
 TATCTTCTAATAAGGTGGTCTGTTTGTCTGAACAAAGTGCTAGACTTTCTGGAGTGATA  
 ATCTGGTGACAAATATTCTCTGTAGCTGTAGCAAGTCACTTAATCTTCTAGCTCTTTT  
 TTCTATCTGCCAAATTGAGATAATGATACTTAACAGTTAGAAGAGGTAGTGTGAATATTA  
 TTAGTTTATATTACTCTTATTCTTTGAACATGAACATATGCCTATGTAGTGTCTTATTGTCT  
 CAGCTGGCTGAGACACGAAAGTCACTGAACAAAACCTACACACGTACCTTCATGTGATT  
 CACTGCCTTCTCTCTACCACTCTATTCCACTGAACAAAACCTACACATACCTTCAT  
 GTGGTTCACTGCTTCTCTCTCTACCAGTCTATTCCACTGAACAAAACCTACGCACATC  
 CTTTCATGTGGCTCAGTGCCTTCTCTCTCTACCAGTCTATTCCATTCTTTCACTGTGCT  
 GACATGTTTGTGCTCTGTTCCATTTTAAACACTGCTCTTACTTTCCAGTCTGTACAGAAT  
 CTATTTCACTTGAGCAAGATGTAAATGGAAGGGTGTGGCAGTGTGCTCTGGAGACTG  
 GATTGAGTCTTGGTGCTATCAATCACCGTCTGTGTTTGTGAGCAAGGCATTTGGCTGCTGTAA  
 GCTTATGCTTCATCTGTAAAGCGGTGTTTGTAAATCTCGATCTTCCCACTCCAGTGTGAT  
 TTGTGGGGATCCAGTGAGATAGAATACATGTAAGTGTGTTTGTAAATTTAAAGATGCTAT  
 ACTAAGGGAAGAAATTGAGGAATTAATGCATACGTTTGGTGTGCTTTTCAATGTTTGA  
 AAATAAAAAAATGTTAAG

10017031.102401

## **FIGURE 98**

></usr/seqdb2/sst/DNA/Dnaseqs.min/ss.DNA52185  
><subunit 1 of 1, 211 aa, 1 stop  
><MW: 22744, pI: 8.51, NX(S/T): 1  
MANAGLQLLGFI LAFLGWIGAIVSTALPQWRIYSYAGDNIVTAQAMYEGLWMSCVSQSTGQI  
QCKVFDSLNLNLSSTLQATRALMVVGILLGVIAIFVATVGMKCMKLEDEDEVQKMRMAVIGGA  
IFLLAGLAILVATAWVGNRIVQEFYDPMTPVNAVYEFQALFTGWAAASLCLLGGALLCCSC  
PRKTTSYPTPRPYPKPAPSSGKDYV

### **Important features:**

#### **Signal peptide:**

amino acids 1-21

#### **Transmembrane domains:**

amino acids 82-102, 118-142 and 161-187

#### **N-glycosylation site.**

amino acids 72-75

#### **PMP-22 / EMP / MP20 family proteins**

amino acids 70-111

#### **ABC-2 type transport system integral membrane protein**

amino acids 119-133

TTCTGGCCAAACCCGGGGCTNCAGCTGTTGGGCTTCATCTCGCCTTCTCTGGGATGGATCGGC  
GCCATCNTCACTGCGCTTCCCCAGTGGAGGATTTTACTCCCTATGCTGGCGACAACATCG  
TGACCGCCAGCCCATGTACGAGGGGCTGTGGATGTCNCGCTGTGCGAGAGCACCGGGCAG  
ATCCAGTGCAAAGTCTTTGACTCCTTGCTGAATCTGAGCAGCACATTGCAAGCAACCCGTGC  
CTTGATGGTGGTTGGCATCCTCCTGGGAGTGATAGCAATCTTTGTGGCCACCGTTGGCATGA  
AGTGTATGAAGTGCTTGAAGACGATGAGGTGCAGAAGATGAGGATGGCTGTCAATTGGGGGC  
GCGATATTCTTCTTGCAAGTCTGGCTATTTTAGTTGCCACAGCATGGTATGGCAATAGAA  
NNTTCAACANTTCTATGACCCCTATGACCCAGTCAATGCCAGGTACGAATTGGTCA  
GGCTCTCTTCACTGGCTGGGCTGCTGCTTCTCTGCGCTTCTGGGAGGTGCCCTACTTTGCT  
GTTCTGTCCC

## FIGURE 100

ACCCTTGACCCAACGCGGCCCCCGACCGNTTCATGGCCAAACGCGGGNCTCCAGCTGTTGG  
GCTTCATTCTCCCCCTCCTGGGATGGACCGGCGCCCATCNTCAGCACTGCCCTGCCCCAGTG  
GAGGATTTACTCCTATNCCGGCNACAACATCGTGACCGCCAGGCCNTGTACGAGGGGCTGT  
GGATGTCTCTGCGTGTGCGCAGAGCACCGGGCAGATCCAGTGCAAAGTCTTTGACTCCCTTGCT  
GAATCTGAGCAGCACATTGCAAGCAACCCGTGCCTTGATGGTGGTTGGCATCCTCCTGGGAG  
TGATAGCAATCTTNNTGGCCACCGTTGTNNNTGAAGTGATGAAGTGCTTGAAGACGATGA  
GGTGCAGAAGATGAGGATGGCTGTCAATTGGGGGCGCGATATTTCTTCTTGAGGTCTGGCTA  
TTTTAGTTGCCACAGCATGGTATGGCAATAGAATCGTTCAAGAATTCTATGACCCATGACCGA

10017081-102401

## FIGURE 101

GGGCCCAGACCATTATCCAACCGGGNTCACTGTTGGGCTCATCTCCCTCCTGGATGAANCGCGC  
CATCNTCAGACTCCCTGCCCCATGGAGATTNNCCTATGCTGGCGACAACATCNTGACCCCC  
AGCCATGTACGAGGGGCTTTGAACGTCNGCGTGTCGCAGANACCGGGCAGATCCAGTGCAA  
AGTCCTTTGACTCCTTGCTGAATCTGNGCAGCACATTGCAGCAACCNTGCCCTGATGGTGGT  
TGGCATCCTCCTGGGAGTGATAGCAATCTTTGTGGCCACCGTTGGCATGAAGTGATGAAGT  
GCTTGAAGACGATGAGGTGCAGAAGATGAGGATGGCTGTTCATTGGGGGCGCGATATTTCTT  
CTTGCAAGTCTGGCTATTTNNNGTTGCCACAGCATGGTATGGCAATAGAATCGTTCAAGAAT  
TCTATGACCCATGACCCAGTCAATGCCAGGTACGAATTTGGTCAGGCTCTCTTCACTGGC  
TGGGCTGCTGCTTCTCTCTGCCTTCTGGGAGGTGCCCTACTTTGCTGTTCTCTGCGA

10017081-102401

## **FIGURE 102**

ATTCTCCCTCCTGGATGGATCGCNCACCGTCACATTGCCTTCCCCANTGGAGGATTNAC  
TCCTATGCTGGCGACAACATCGTGACCCCCAGGCCATTTACCGAGGGGCTTGGATGTCNT  
GCNTGTGCGAGAGCACCGGGCAGATCCAGTGCAAAGTCTTTGACTCCTTGCTGAATCTGAG  
CAGCACATTGCAAGCAACCCGTGCCTTGATGGGGTTGGCATCCTCCTGGGAGTGATAGCAAC  
CTTTGTGGCCACCGTTGGCATGAAGTGTATGAAGTGCTTGAAGACGATGAGGTGCCAGAAG  
ATGAGGATGGCTGT CAT TGGGGGCGCGATATTTCTTGTTCAGGTCTGGCTATTTTAGTNGC  
CACAGCATGGTATGGCAATAGANTNNTTCNNGNNNTCTATGACCCATGACCCAGTCAATG  
CCAGGTACGAATTTGGTCAGGCTCTCTTCACTGGCTGGGCTGCTGCTTCTCTGCTTCTG  
GGAGGTGCCCTACTTTGCTGTTCTCTGTCCC

## **FIGURE 103**

AGAGCACCGGCAGATCCCAGTNCAAAGTCTTTGACCCTTGCTGAATCTGAGCAGCACATTNC  
AAGCAACCCCTTGCCCTGAAGGTGGTTGNCATCCCCCTGGGAGTGAATAGCAATCTTTGTG  
GCCACCGTTGGCATGAAGTNTATGAAGTGCTTGGAAGACGATGAGGTGCAGAAGATGAGGAT  
GGCTGTCAATTGGGGGCGCGATATTTCTTCTTGACAGTCTGGCTATTTTAGTNNCCACAGCAT  
GGTATGGCAATAGNATNNTTTCGNGGNTTCTATGACCCTATGACCCAGTCAATGCCAGGTAC  
GAATTGGTCAGGCTCTCTTCACTGGCTGGGCTGCTGCTTCTCTCTGCCTTCTGGGAGGTGC  
CCTACTTTGCTGTTCTCTGTCCCGAA

10017081-102401

## **FIGURE 104**

AGCAATGCCCTGCCCCAGTGGAGGATTAATTCCTATGNTGGGGACAACATTGTGACNGCCC  
AGGCCATGTACGGGGGGCTGTGGATGTCTGCGTGTGCGAGAGCACCGGGCAGATCCAGTGC  
AAAGTNTTTGACTCCTTGCTGAATTTGAGCAGCACATTGCAAGCAACCCGTGCCTTGATGGT  
GGTTGGCATCTTCCTGGGAGTGATAGCAATCTTTGTGGCCACCGTGGNAATGAAGTGTATGA  
AGTGCTTGGAAGACGATGAGGTGCAGAAGATGAGGATGGCTGTCAATTGGGGGCGCGATATTT  
CTTNTTGCAGGTCTGGCTATTTTAGTTGCCACAGCATGGTATGGCAATAGAATNGTTCAGA  
ATTTTATGACCCATGACCCAGTCAATGCCAGGTACGAATTTGGTCAGGCTTTNTTCACTG  
GCTGGGCTGCTGCTTNTTCTGCCTTNTGGGAGGTGCCCTANTTTGCTGTTCTCTGCGAACC

101/031-102409

## **FIGURE 105**

TCATAGGGGGCGCGATATTTTTTCTTGCAGGTNTGGTTATTTTAGTTGCCACAGCATGGTA  
TGGCAATAGAATCGTTCAAGAATTNTATGACCCTATGACCCAGTCAATGCCAGGTACGAAT  
TTGGTCAGGCTCTNTTCACTGGNTGGGCTGCTGCTTCTNTNNGCCTTNTGGGAGGTGCCCTA  
CTTTGCTGTTCCIG

104201-180/1001

## **FIGURE 106**

TTCTCTGGGATGGATCCGCCCCCATCNTCACATGCCCTGCCCCNTGGAGATTTACNCCTATGC  
TGGCGAACACATCNTGACCGCCCAGGCCATGTACGAGGGCTGTGGAATGTCCTGCGTGTC  
CCAGAGCACCGGCAGATCCAGTGC AAAGTCTTTGACTCCTTGCTGAATCTGAGCAGCACAT  
TGCAAGCAACCNTGCC TTGATGGTGGTTGGCATCCTCCTGGGAGTGATAGCAATCTTTGTGG  
CCACCGTTGGCATGAAAGTGATGAAGTGCTTGGAAGACGATGAGGTGCAGAAGATGAGGAT  
GGCTGTCATTGGGGGCGGATATTTCTTCTTG CAGGTCTGGCTATTTTAGNNGCCACAGCAT  
GGTATGGCAATCAGACCCNNTCANAACTCTATGACCCCTATGACCCCAGTCAATGCCAGGTA  
CGAATTTGGTCAGGCTCTCTTCACTGGCTGGGCTGCTGCTTCTCTGCTTCTGGGAGGTG  
CCCTACTTTGCTGTTCTGTGCCCCGAAAAACAACCTCTTACCCACG

10017081-10240

## **FIGURE 107**

CGGGGCTGCAGCTGTTGGGCTTCATCTCGCTTCCTGGGATGGAATCGGCGCCATCGTCAGCA  
CTGCCCTGCCCCATGGAGGATTTACTCNTATGCTGGCGACAACATCGTGACCNCCCAGGCCA  
TGTACGAGGGGCTGTGGATGTCNCGTGTCGCAGAGCACCGGGCAGATCCAGTGCAAAGTCT  
TTGACTCCTTGCTGAATCTGAGCAGCACATTGCAAGCAACNTGCCTTGATGGTGGTTGGCA  
TCCTCCTGGGAGTGATAGCAATCTTTGTGGCCACCGTTGGCATGAAGTGTATGAAGTGCTTG  
GAAGACGATGAGGTGCAGAAGATGAGGATGGCTGTCAATGGGGGCGCGATATTTCTTCTTGC  
AGGTCTGGCTATTTINTAGTTGCCACAGCATGGTATGGCAATAGAATCGTTCAAGAATTCTAT  
GACCCTATGACCCAGTCAATGCCAGGTACGAATTTGGTCAGGCTCTCTTCACTGGCTGGGC  
TGCTGCTTCTCTCTGCCTTCTGGGAGGTGCCCTACTTTGCTGTTCCTGCGAA

107081-102401

# THEORY

GCGTGCCTCAGCTCGCGGGGACCGCGGGCCCGCCCTCGCCCTCCGCCCTCGCCCTGCA  
CGCGTAGACACGCCACCCCTCTCCAGCGCGCCACCGCGGTAGAGGACCCCGCCCGCTGCCCG  
ACCGGTCCCCGCTTTTTGTAAACCTTAAGCGGGCGAGCATTAACGCTTCCCGCCCGGT  
GACCTCTCAGGGGTCTCCCGGCCAAAGGTGCTCCGCCGTAAAGGAACATGGCGAAGGTGGAG  
CAGGTCTGAGCGCTCGAGCGCGACAGCAGCTCAAATTCGAGGCTCCCTCCACCGATGTTGT  
CACACCAACCTAAAGCTTGGCAACCCGACAGACCGAAATGTGTGTTTAAAGTGAAGATCA  
CAGCACACGTTAGGTACTGTGTGAGGCCCAACAGCGGAATCATCGATGCAGGGGCTCAAT  
AATGTATCTGTGATGTTACAGCCTTTCGATATGATATGATCCCAATGGAGAAAATTAACACAGTT  
TATGGTTCAGTCTATGTTTGTCTCCAATGCACATCTCAGATATGGAAGCAGTATGAAGAGG  
CAAAACCGGAAGACCTTATGGAATCAAAACTTAGATGTGTGTTTGAATTGCCAGCAGAGAA  
GATAAACACCATGATGTAGAATAATAAATAATATATCCCAACTGCATCAAGACAGAAAC  
ACCAATAGTGTCTAAGTCTCTGAGTCTCTTTGGATGACACCGGAAGTTAAGAGGTTATGG  
AAGAAGTGAAGAGGCTGCAGGTGAAGTTCAGAGGCTACGGGAGGAGAACAGCAGGTTCAAG  
GAAGAAGATGGAGTCGGATGAGGAAGCAGTGCAGAGCAACAGGCCCTTTCAGATTAGC  
CCCAACTGGGAAGGAAGAGGCCCTTAGCACCGGCTCTGGCTCTGGTGGTTTTGTCTTTA  
TCGTTGGTGTAAATTTATGGGAAGATTTGCCCTGTGAGAGGTAGCATGCAAGGATGTAAATTG  
GATTGGTGGATCCACCATATCATGGGATTTAAATTTATCAAAACATGTGTAAAGAATAAT  
AATGTATGATGACATCTACAGGCTTTCGCTTTAAATACCTCCCTCGCACACATACAC  
AGATACACACACAAATATATATGATCAAGCATCTTTAGAAAGGTAAAAATGTATAGTAAC  
ATTGAGGGGGAAAAAGATGATCTTTATTAATGACAAGGGAAACCATGAGTATATGCCAAAT  
GGCATATTGTAATGTCAATTTTAAACATTGGTAGGCCCTTGGTACATGATGCTGGATTACCTC  
TCTTAAATATGACACCCCTCTCGCTGTGGTGTGCGCCCTTGGGAGCTGGAGGCCAGGT  
CTCTGGGAGTGCGGTCACTCTCACACAGTAGTCCCAAGCTGGCCCATCCCGGCCACGCTG  
CTTCCGTGCTCTCAGTCTGTCCAAGCCATCAGCTCCTTGGGACTGATGAACAGAGTCAGA  
AGCCCAAAGGAATTTGCATCTGTGGCAGCATCAGCATCTCGTCATAAGTGAAGCGGTGTGT  
TGACTATGTGACCGCGCTTTGGAAATAATGAGCAGTGCTTTTGTTCACCTAAAGGGACAA  
GCTAAATTTGATATGGTTTCATGTAGTGAAGTCAAATGCTTTATTCAGAGGTTTTATGCATA  
TTTAACTTATTAATGTTATTTATCATCTCATGTTTCTTATTGTCAGAGAGTACAGTTAATGC  
TGCGTGCTGCTGAATCTGTTGGGTGAATGTTATGCTGCTGGAGGGCTGGGCTCCTCT  
GTCTCTGGAGAGTCTGTGTCATGTGAGGAGTGGGGTTTATTGGGATCTGGGAGAGAGCTGCA  
GGAAGTGTTTTTCTGGCTAGTAATAACAACCTGTCATAGGAGGAGAAATCTCAGTAGTG  
ACAGTCAACTCTAGGTTACCTTTTTTAATGAAGAGTAGTCAGTCTTCTAGATTGTTCTTATA  
CCACTCTCAACACTTACATCACTTCACGCGCCAGCTGCCAGTCTGAGCTGACCTCCCC  
TTGGGACCTAGCTCGGAGTCAGGACCAATGATTCGGCTGTCAGAGGTTAAGACGAGGCG  
ACCAGCAGTTGTGGGTGGGAGCAAGGGAAGAGAGAACTCTTCAGCGAATCCTTCTAGTAC  
TAGTTGAGAGTCTGACTGCTGAATTAATTTATGCTACATAAAGAACCAACCTGTTCTGTTGA  
CTATGTAGCATCTTTGAAAAGAAAATAATTAATAAGCTCCAAAATTAAGAAA

## **FIGURE 109**

</usr/seqdb2/sst/DNA/Dnaseqs.min/ss.DNA53977

<subunit 1 of 1, 243 aa, 1 stop

<MW: 27228, pI: 7.43, NX(S/T): 2

MAKVEQVLSLEPQHCLKFRGPFTDVVTNLKLGNPIDRNVCFKVKTAPRRYCVRENSGIID  
AGASINVSVMQLQPFDDYDPNEKSKHKFMVQSMFAPTDTSMEAVWKEAKPEDLMDSKLRCVFE  
LPAENDKPHDVEINKIIISTTASKTETPIVSKSLSSSLDDTEVKKVMEECKRLQGEVQRLREE  
NKQFKEEDGLRMRKTVQSNPISALAPTGKEEGLSTRLLALVVLFFIVGVIIGKIAL

**Important features:**

**Transmembrane domain:**

amino acids 224-239

**N-glycosylation site.**

amino acids 68-71

**N-myristoylation site.**

amino acids 59-64, 64-69 and 235-240

## **FIGURE 110**

GTCAGTCTTCTAGATTGTCCTTATCCACCTTTCAACCANTACTCACATTTNAGCGCCCAG  
GTCCANGTCTGAGCCTGACTTCCCCTTGGGGACCTAGCCTGGAGTCAGGACAATGGNTCGGG  
CTGCAGAGGNTTAGAAGCGAGGGCACCAGCAGTTTGGGTGGGAGCAAGGNNGAGAGAAA  
CTCTTCAGCGAATCCTTCTAGTACTAGTTGAGAGTTTGACTGTGAATTAATTTTATGCCATA  
AAAGACNAACCCAGTTCTGTTTGACTATGTAGCATCTTGAAAAGAAAAATTATAATAAAGCC  
CCAAAATTAAGAATTCTTTTGTCATTTTGTCACATTTGCTCTATGGGGGGAATTATTATTTT  
ATCATTTTTATTATTTTGCCATTGGAAGGTTAACTTTAAAATGAGC

bioRxiv preprint doi: <https://doi.org/10.1101/000000>; this version posted January 1, 2015. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

## **FIGURE 111**

TATTGTAAAGGCCATTTTAAACCATTGGTAGGCCTTGGTACATGATGCTGGATTACCTCCTT  
AAATGACACCNNTTCCTCGCCTGTTGGTGCTGGCCNTTGGGGAGCTGGAGCCCAGCATGCTG  
GGGAGTGCGGTCAGCTCCACACAGTAGTCCCCACGTGGCCCACTCCCGGCCAGGCTGCTTT  
CCGTGTCTTCAGTTCTGTCCAAGCCATCAGCTCCTTGGGACTGATGAACAGAGTCAGAAGCC  
CAAAGGAATTGCCACTGTGGCAGCATCAGACGTACTCGTCATAAGTGAGAGGCGTGTGTTGA  
CTGATTGACCCAGCGCTTTGGAAATAAATGGCAGTGCTTTGTTCACTTAAAGGGACCAAGCT  
AAATTGTATTGGTTCATGTAGTGAAGTCAAACGTATTATTCAGAGATGTTTAATGCATATTTA  
ACTTATTTAATGTATTTTCATCTCATGTTTTCTTATTGTCACAAGAGTACAGTTAATGCTGCG  
TGCTGCTGAACTCTGTTGGGTGAACTGGTATTGCTGCTGGAGGGCTG

1007031-102401

	1980	1981	1982	1983	1984	1985	1986	1987	1988	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035	2036	2037	2038	2039	2040	2041	2042	2043	2044	2045	2046	2047	2048	2049	2050	2051	2052	2053	2054	2055	2056	2057	2058	2059	2060	2061	2062	2063	2064	2065	2066	2067	2068	2069	2070	2071	2072	2073	2074	2075	2076	2077	2078	2079	2080	2081	2082	2083	2084	2085	2086	2087	2088	2089	2090	2091	2092	2093	2094	2095	2096	2097	2098	2099	2100	2101	2102	2103	2104	2105	2106	2107	2108	2109	2110	2111	2112	2113	2114	2115	2116	2117	2118	2119	2120	2121	2122	2123	2124	2125	2126	2127	2128	2129	2130	2131	2132	2133	2134	2135	2136	2137	2138	2139	2140	2141	2142	2143	2144	2145	2146	2147	2148	2149	2150	2151	2152	2153	2154	2155	2156	2157	2158	2159	2160	2161	2162	2163	2164	2165	2166	2167	2168	2169	2170	2171	2172	2173	2174	2175	2176	2177	2178	2179	2180	2181	2182	2183	2184	2185	2186	2187	2188	2189	2190	2191	2192	2193	2194	2195	2196	2197	2198	2199	2200	2201	2202	2203	2204	2205	2206	2207	2208	2209	2210	2211	2212	2213	2214	2215	2216	2217	2218	2219	2220	2221	2222	2223	2224	2225	2226	2227	2228	2229	2230	2231	2232	2233	2234	2235	2236	2237	2238	2239	2240	2241	2242	2243	2244	2245	2246	2247	2248	2249	2250	2251	2252	2253	2254	2255	2256	2257	2258	2259	2260	2261	2262	2263	2264	2265	2266	2267	2268	2269	2270	2271	2272	2273	2274	2275	2276	2277	2278	2279	2280	2281	2282	2283	2284	2285	2286	2287	2288	2289	2290	2291	2292	2293	2294	2295	2296	2297	2298	2299	2300	2301	2302	2303	2304	2305	2306	2307	2308	2309	2310	2311	2312	2313	2314	2315	2316	2317	2318	2319	2320	2321	2322	2323	2324	2325	2326	2327	2328	2329	2330	2331	2332	2333	2334	2335	2336	2337	2338	2339	2340	2341	2342	2343	2344	2345	2346	2347	2348	2349	2350	2351	2352	2353	2354	2355	2356	2357	2358	2359	2360	2361	2362	2363	2364	2365	2366	2367	2368	2369	2370	2371	2372	2373	2374	2375	2376	2377	2378	2379	2380	2381	2382	2383	2384	2385	2386	2387	2388	2389	2390	2391	2392	2393	2394	2395	2396	2397	2398	2399	2400	2401	2402	2403	2404	2405	2406	2407	2408	2409	2410	2411	2412	2413	2414	2415	2416	2417	2418	2419	2420	2421	2422	2423	2424	2425	2426	2427	2428	2429	2430	2431	2432	2
--	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	---

CCCTGGTGGTTTTGTCTCTTAATTGTTGGTGTAATTNTTGGGAAGATTGCTTGTAGAGGTA  
GNATGCACNCGGCTGGTAAATTGGATTGGTGGATCCACATATCCATGGGATTTAAATTAT  
CATAACCATGTGTAAAAAGAAATTATGTATGATGACATNTCACAGGTATTGCCTTTAAATT  
ACCCATCCCTGNANACACATACACAGATACACANANACAAATNTAATGTAAACGATNTTTTAG  
AAAGTTAAAAATGTATAGTAAC

## FIGURE 113

GGTGGCCCATTCCTGGCCAGGCTGCTTTCCGGTNTTCAGTTCTGTCCAAGCCATCAGCTCC  
TTGGGACTGATGAACAGAGTCAGAAGCCCAAAGGAATTGCAC TGTGGCAGCATNAGACGTAC  
TTGTNATAAGTGAGAGGCGTGTGTTGACTGATTGACCCAGCGCTTTGGAAAATAAATGGCAGT  
GCTTTGTTCAANTTAAAGGGACCAAGCTAAATTTGTATTGGTTCATGTAGTGAAGTCAAAC TG  
TTATTCAGAGATGTTTAAATGCATATTTAANTTATTTAATGTATTTNATNTCATGTTTTCTTA  
TTGT CACAAGAGTACAGTTAATGCTGCGTGCTGCTGAANTNTGTTGGGTGAACTGGTATTGC  
TGCTGGAGGGCTGTGGGCTCCTCTGTCTTTGGAGAGTCTGGTCATGTGGAGGTGGG

bioRxiv preprint doi: <https://doi.org/10.1101/000000>; this version posted January 1, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

## **FIGURE 114**

TGCTTTCCGTGTCTTCAGTTCTGTCCAAGCCATCAGCTCCTTGGGACTTGATGAACAGAGTC  
AGAAGCCCAAAGGAATTGCACTGTGGCAGCATCAGACGTACTCGTCATAAGTGAGAGGCGTG  
TGTTGACTGATTGACCCAGCGCTTTGGAAATAAATGGCAGTGCTTTGTTCACTTAAAGGGAC  
CAAGCTAAATTTGTATTGGTTCATGTAGTGAAGTCAAAC TGTATTTCAGAGATGTTTAATGC  
ATATTTAACTTATTTAATGTATTTTCATCTCATGTTTTCTTATTGTCAAGAGTACAGTTAA  
TGCTGCGTGTC

10017081-102401

## **FIGURE 115**

AAACCTTTAAAAGTTGAGGGGAAAAGAATGATCCTTTATTAATGACAAGGGAAACNTGNGT  
AATGCCACAATGGCATATTGTAAATGTCATTTTAAACATTGGTAGGCCTTGGTACATGATGC  
TGGATTACCTCTCTTAAAATGACACCCTTCCTCGCCTGTTGGTGCTGGCCCTTGGGGAGCTN  
GAGCCCAGCATGCTGGGGAGTGCGGTCTGCTCCACACAGTAGTCCCCANGTGGCCCANCCCC  
GGCCCAGGCTGCTTTCCGTGTCCTTCAGTTCTGTCCAAGCCATCAGCTCCTTGGGANTGATGA  
ACAGAGTCAGAAGCCCAAAGGAATGTCANTGTGGCAGCATCAGANGTANTNGTCATAAGTGA  
GAGGCGTGTGTTGANTGATTGACCCAGCGCTTTGGAAATAAATGGCAGTGCTTTGTTTCANTT  
AAAGGGNCCAAGNTAAATTTGTATTGGTTCATGTAGTGAAGTCAAANTGTTATTCAGAGATG  
TTTAATGCATATTTAANTTATTTAATGTATTTTCATNTCATGTTTCTTATTGTACAAGGGT  
ACAGTTAATGCTGCTGCTGCTGAANTCTGTTGGGTGAANTGGTATTGCTG

10017051-102401

## **FIGURE 116**

GGCCCTTGGGGAGCTGGAGCCCAGCATGCTGGGGAGTGCGGTCAGCTCCACACAGTAGTCCC  
CACGTGGCCCACTCCCGGCCCAGGCTGCTTTCCGTGTCTTCAGTTCGTCCAAGCCATCAGC  
TCCTTGGGACTGATGAACAGAGTCAGAAGCCCAAAGGAATTGCACTGTGGCAGCATCAGACG  
TACTCGTCATAAGTGAGAGGCGTGTGTGACTGATTGACCCAGCGCTTGGAAATAAATGGC  
AGTGCTTTGTTCACCTTAAAGGGACCAAGCTAAATTGTATTGGTTCATGTAAGTGAAGTCAA  
CTGTATTTCAGAGATGTTAATGCATATTAACTTATTTAATGTATTTTCATCTCATGTTTTTC  
TTATTGTCACAAGAGTACAGTTAATGCTGCGTGCTGCTGAACTCTGTTGGGTGAACTGGTAT  
TGCTGCTGGAGGGCTGTGGGCTCCTCTGTCTCTGGAGAGTCTGGTCATGTGGAGGTGGG

GGCCCTTGGGGAGCTGGAGCCCAGCATGCTGGGGAGTGCGGTCAGCTCCACACAGTAGTCCC  
CACGTGGCCCACTCCCGGCCCAGGCTGCTTTCCGTGTCTTCAGTTCGTCCAAGCCATCAGC  
TCCTTGGGACTGATGAACAGAGTCAGAAGCCCAAAGGAATTGCACTGTGGCAGCATCAGACG  
TACTCGTCATAAGTGAGAGGCGTGTGTGACTGATTGACCCAGCGCTTGGAAATAAATGGC  
AGTGCTTTGTTCACCTTAAAGGGACCAAGCTAAATTGTATTGGTTCATGTAAGTGAAGTCAA  
CTGTATTTCAGAGATGTTAATGCATATTAACTTATTTAATGTATTTTCATCTCATGTTTTTC  
TTATTGTCACAAGAGTACAGTTAATGCTGCGTGCTGCTGAACTCTGTTGGGTGAACTGGTAT  
TGCTGCTGGAGGGCTGTGGGCTCCTCTGTCTCTGGAGAGTCTGGTCATGTGGAGGTGGG

# FIGURE 117

GCGAGCTCCGGGTGCTGTGGCCCCGGCCTTGGCGGGGGCGGCCTCCGGCTCAGGCTGGCTGAGA  
 GGCCTCCAGCTGCAGCGTCCCCGCCCGCCTCCTCGGGAGCTCTGATCTCAGCTGACAGTGCC  
 CTCGGGGACCAAAACAGCCTGGCAGGGTCTCACITTTGTTGCCAGGCTGGAGTTTCAGTGCCA  
 TGATCATGGTTTACTGCAGCCTTGACCTCTCGGGTTCAAGCGATCCTGCTGAGTAGCTGGGA  
 CTACAGGACAAAATTAGAAGATCAAAATGGAAAATATGCTGCTTTGGTTGATATTTTCCACC  
 CCTGGGTGGACCTCATTGATGGATCTGAAATGGAATGGGATTTTATGTGGCACTTGAGAAA  
 GGTACCCCGGATTTGTCAGTGAAGGACATTTCCATCTCACCAGCCCCGATTTGAGGCAGATG  
 CTAAGATGATGGTAAATACAGTGTGTGGCATCGAATGCCAGAAAGAACTCCCAACTCCCGAGC  
 CTTTCTGAATTTGGAGGATTATCTTTCTATGAGACTGTCTTTGAGAATGGCACCCGAACCTT  
 AACCAGGGTGAAGATTCAAGATTTGGTTCTTGAGCCGACTCAAAATATCACCACAAAGGGAG  
 TATCTGTTAGGAGAAAGAGACAGGTGTATGGCACGACAGCAGGTTTTCAGCATCTTGGACAAA  
 AGGTTCTTAACCAATTTCCCTTTTCAGCACAGCTGTGAAGCTTTCCACGGGCTGTAGTGGCAT  
 TCTCATTTCCCTCAGCATGTTCTAATGCTGCCCACTGTGTTTCATGATGAAAGGACTATG  
 TCAAAGGGAGTAAAAAGCTAAGGGTAGGGTTTGTGAAGATGAGGAATAAAAGTGGAGGCAAG  
 AAACGTTCGAGGTTCTAAGAGGAGCAGGAGAGAAGCTAGTGGTGGTGACCAAGAGAGGGTAC  
 CAGAGAGCATCTGCAGGAGAGACGCAAGGGTGGGAGAGAAGAAAATCTGGCGGGGCTC  
 AGAGGATTGCCGAAGGGAGGCCCTTCTTTTCAGTGGACCCGGGTCAAGAATACCCACATTCGG  
 AAGGGCTGGGCACGAGGAGGCATGGGGGACGCTACCTTGGACTATGACTATGCTCTTCTGGA  
 GCTGAAGCGTGTCTACAAAAGAAATACATGGAACCTTGAATCAGCCCAACGATCAAGAAAA  
 TGCCCTGGTGGAAATGATCCACTTCTCAGGATTTGATAACGATAGGGCTGATCAGTTGGTCTAT  
 CGGTTTTGCAGTGTGTCCGACGAATCCAATGATCTCCTTTACCAATACTGCGATGCTGAGTCT  
 GGGCTCCACCGGTTCCGGGGTCTATCTGCGTCTGAAAGATCCAGACAAAAGAATTGGAAGC  
 GCAAAATCATTGCGGTCTACTCAGGGCACAGTGGGTGGATGTCCACGGGGTTCAGAAGGAC  
 TACAACGTTGCTGTTTCGCATCACTCCCTAAAATACGCCAGATTGCTCTGGAATTCACGG  
 GAACGATGCCAATTTGTCTTACGGCTAAACAGAGACCTGAAACAGGGCGGTGTATCATCTAAA  
 TCACAGAGAAAACCGCTCTGCTTACCGTAGTGAGATCACTTCATAGGTTATGCTTGGACTT  
 AAACCTTCAAAACAGGTATTATAAATAACATGTGACTCCTTAATGGACTTATTTTCAGGGTCC  
 TACTCTAGAAGAAATCTAATAGGATGCTGGTTGTGTATTAAATGTGAAATTGCATAGATAAA  
 GGTAGATGGTAAAGCAATTAGTATCAGAATAGAGACAGAAAGTTACAACAGGTTTGTACTA  
 CTCTGAGATGGATCCATTTCAGCTCATGCCCTCAATGTTTATATTGTGTATCTGTTGGGTCT  
 GGGACATTTAGTTTAGTTTTTTTGAAGAATTACAAATCAGAAGAAAAAGCAAGCATTATAAA  
 CAAAACATAAAGCTGTTTACTGCTTTAAGAAATAACAATTACAAATGTGTATTATTGTTAAAAA  
 TGGGAGAAATAGTTTGTCTATGAAATAAACCTAGTTTGAAGAAAGGGAAGCTGAGACATTT  
 TAAGATCTCAAGTTTTTATTAACATAACTCAAAATATGGACTTTTCATGTATGTCATAGGG  
 AAGACATTCACAAATATGAATGATCATGTGTGAAAGCCACATTATTTTATGCTATACAT  
 TCTATGTATGAGGTGCTACATTTTATGAGCAAAAGAATTCTGTAATCTTTTCAAGAAAGAGT  
 CTTTTTCTCCTTGACAAAATCCAGCTTTTGTATGAGGACTATAGGGTGAATCTCTGATTAG  
 TAATTTTAGATATGCTCTTCTCTAAAATGAATAAAATTTATGAATATGA

## **FIGURE 118**

```
</usr/seqdb2/sst/DNA/Dnaseqs.min/ss.DNA57253
<subunit 1 of 1, 413 aa, 1 stop
<MW: 47070, pI: 9.92, NX(S/T): 3
MENMLLWLIFFTPGLIDGSEMEWDFMWHLRKVPRIVSERTFHLTSPAFEADAKMMVNTVC
GIECQKELPTPSLSLEEDYLSYETVVFENGTRTLTRVKVQDLVLEPTQNITTKGVSVRRKRQV
YGTDSRFSILDKRFLTNFPFSTAVKLSTGCSGILISPQHVLTAACVHDGKDYVKGSKKLRV
GLLKMNRKSGGKKRRGSKRSRREASGGDQREGTREHLQERAKGGRRRKSGRGQRIAEGRPS
FQWTRVKNTHIPKGWARGMGDATLDYDYALLELKRAHKKKYMELGISPTIKKMPGGMIHFS
GFDNDRADQLVYRFCSVSDSNDLLYQYCDAESGSGVYLRLKDPDKKNWKRKIIAVYSG
HQWVDVHGVQKDYNAVAVRITPLKYAQICLWIHGNDANCAYG
```

### **Important features:**

#### **Signal peptide:**

amino acids 1-16

#### **N-glycosylation sites.**

amino acids 90-93, 110-113 and 193-196

#### **Glycosaminoglycan attachment site.**

amino acids 236-239

#### **Serine proteases, trypsin family, histidine active site.**

amino acids 165-170

## FIGURE 119

AATGTGAGAGGGGCTGATGGAAGCTGATAGGCAGGACTGGAGTGTTAGCACCAGTACTGGAT  
GTGACAGCAGGCAGAGGAGCACTTAGCAGCTTATTTCAGTGTCCGATTCTGATTCGGCAAGG  
ATCCAAGCATGGAATGCTGCCGTCGGGCAACTCCTGGCACACTGCTCCTCTTTCTGGCTTTT  
CTGCTCCTGAGTTCAGGACCGCACGCTCCGAGGAGGACCGGGACGGCCTATGGGATGCCTG  
GGGCCCATGGAGTGAAATGCTCACGCACCTGCGGGGAGGGGCCCTCTACTCTCTGAGGCGCT  
GCCCTGAGCAGCAAGAGCTGTGAAGGAAGAAATATCCGATACAGAACATGCAGTAATGTGGAC  
TGCCCCACCAGAAGCAGGTGATTTCCGAGCTCAGCAATGCTCAGCTCATAATGATGTCAAGCA  
CCATGGCCAGTTTTATGAATGGCTTCCTGTGTCTAATGACCCCTGACAACCCATGTTCACTCA  
AGTGCCAAGCCAAAGGAACAACCCCTGGTTGTTGAACTAGCACCTAAGGTCCTTAGATGGTACG  
CGTTGCTATACAGAATCTTTGGATATGTGCATCAGTGGTTTATGCCAAATTGTTGGCTCGCA  
TCACCAGCTGGGAAGCACCGCTCAAGGAAGATAACTGTGGGCTGTGCAACGGAGATGGGTCCA  
CCTGCCGGCTGGTCCGAGGGCAGTATAAATCCCAGCTCTCCGCAACCAAACTCGGATGATACT  
GTGTTGTGCACTTCCCTATGGAAGTAGACATATTGCGCTTGTCTTAAAGGTCCTGATCACTT  
ATATCTGGAACCAAAAACCCCTCCAGGGGACTAAAGGTGAAACAGTCTCAGCTCCACAGGAA  
CTTTCCTTGTGGACAATTCTAGTGTGGACTTCCAGAAATTTCCAGACAAAGAGATACTGAGA  
ATGCTGGACCACTCACAGCAGATTTCAATTGTCAAGATTCTGAACCTCGGGCTCCGCTGACAG  
TACAGTCCAGTTCATCTTCTATCAACCCATCATCCCGATGGAGGGAGACGGATTTCTTTT  
CTTGCTCAGCAACCTGTGGAGGAGGTTATCAGCTGACATCGGCTGAGTGCTACGATCTGAGG  
AGCAACCGTGTGGTTGCTGACCAATACTGTCACTATTACCCAGAGAACATCAAACCCAAACC  
CAAGCTTCAGGAGTGCAACTTGGATCCTTGTCCAGCCAGTGACGGATACAAGCAGATCATGC  
CTTATGACCTCTACCATCCCTTCTCCTCGGTGGGAGGCCACCCCATGGACCGCGTGTCTCTCC  
TCGTTGAGGGGGGGCATCCAGAGCCGGGCAGTTTCTGTGTGGAGGAGGACATCCAGGGGCA  
TGTCACTTCAGTGGAAGAGTGGAATGCATGTACACCCCTAAGATGCCCATCGCGCAGCCCT  
GCAACATTTTTGACTGCCCTAAATGGCTGGCACAGGAGTGGTCTCCGTGCACAGTGACATGT  
GGCCAGGGCCCTCAGATACCGTGTGGTCTCTGCATCGACCATCGAGGAATGCACACAGGAGG  
CTGTAGCCCAAAAAACAAGCCCCACATAAAAGAGGAATGCATCGTACCCACTCCCTGCTATA  
AACCCAAAGAGAAACTTCCAGTCGAGGCCAAGTTGCCATGGTTCAAACAAGCTCAAGAGCTA  
GAAGAAGGAGCTGCTGTGTGTCAGAGGAGCCCTCGTAAAGTTGTAAAGCACAGACTGTTCTATA  
TTTGAACACTGTTTGTTTAAAGAAAGCAGTGTCTCACTGGTTGTAGCTTTTCATGGGTTCTGA  
ACTAAGTGAATCATCTCACCAAGCTTTTGGCTCTCAAATTAAGATTGATTAGTTTCAA  
AAAAAAAA

## **FIGURE 120**

</usr/seqdb2/sst/DNA/Dnaseqs.min/ss.DNA58847  
<subunit 1 of 1, 525 aa, 1 stop  
<MW: 58416, pI: 6.62, NX(S/T): 1  
MECCRRATPGTLLLFLAFLLLSSRTARSEEDRDGLWDAWGPWSECSRTC GGGASYS LRRLCS  
SKSCBGRNIRYRTC SNVDCPPEAGDFRAQQCSAHNDVKHHGQFYEWLPVSNPDNP CSLKCO  
AKGTTLVVELAPKVL DGTCTYTESLDMCISGLCQIVGCDHQLGSTVKEDNCGVCNGDGSTCR  
LVRGQYKSQLSATKSDDTVVALPYGSRHIRLV LKGPDLHLYLETKTLQGTKGENSL SSGTFL  
VDNSSVDFQKFPDKEILRMAGPLTADFIVKIRNSGSADSTVQFIFYQPIIHRWRETDFFPCS  
ATCGGGYQLTSAECYDLRSNRVVADQYCHYYPENIKPKPKLQECNLDP CPASDGYKQIMPYD  
LYHPLPRWEATPWTACSSSCGGGIQSRVSCVEEDIQGHVTSVEEWKCMYTPKMPIAQPCNI  
FDCPKWLAQEWSPCTVT CGQGLRYRVVLCIDHRGMHTGGCSPKTKPHIKEECIVPTPCYKPK  
EKLPVEAKLPWFKQAQELEEGAAVSEEPS

### **Important features:**

#### **Signal peptide:**

amino acids 1-25

#### **N-glycosylation site.**

amino acids 251-254

#### **Thrombospondin 1**

amino acids 385-399

#### **von Willebrand factor type C domain proteins**

amino acids 385-399, 445-459 and 42-56

## **FIGURE 121**

CGGACGCGTGGGCGGCGGCTGCGGAACTCCCGTGGAGGGGCCGTGGGCCCTCGGGCCTGAC  
AGATGGCAGTGGCCACTGCGGCGGCAGTACTGGCCGCTCTGGGCGGGGCGCTGTGGCTGGCG  
GCCGCCGGTTCTGTGGGGCCAGGGTCCAGCGGCTGCGCAGAGGCGGGGACCCCGGCCCTCAT  
GCACGGGAAGACTGTGTGATCACCGGGCGAACAGCGGCTGGGCCGCGCCACGGCCGCCG  
AGCTACTGCGCCTGGGAGCGCGGGTGATCATGGGTGCGCGGACCGCGCGCGCCGAGGAG  
GCGGCGGGTCAGTCCGCCCGAGCTCCGCCAGGCCGCGGAGTGGGCCCGAGACCTTGGCGT  
CAGCGGGGTGGGCGAGCTCATAGTCCGGGAGCTGGACCTCGCCTCGTGCCTCGGTGCGCG  
CCTTCTGCCAGGAAATGCTCCAGGAAGAGCCTAGGCTGGATGTCTTGATCAATAACGCAGGG  
ATCTTCCAGTGCCCTTACATGAAGACTGAAGATGGGTTTGAGATGCAGTTCGGAGTGAACCA  
TCTGGGGCACTTTCTACTCACCAATCTTCTCCTTGGACTCCTCAAAAGTTCAGCTCCCAGCA  
GGATTGTGGTAGTTTCTTCAAACCTTTATAAATACGGAGACATCAATTTTGATGACTTGAAC  
AGTGAACAAAGCTATAATAAAAGCTTTTGTATAGCCGAGCAAACCTGGCTAACATTCTTTT  
TACCAGGGAAGTAGCCCGCCGCTTAGAAGGCACAAATGTACCGTCAATGTGTGTCATCCTG  
GTATTGTACGGACAAATCTGGGGAGGCACATACACATTCCACTGTTGGTCAAACCACTCTTC  
AATTTGGTGTCAATGGGCTTTTTTCAAACCTCCAGTAGAAGGTGCCCAGACTTCCATTTATTT  
GGCCTCTTCACTGAGGTAGAAGGAGTGTGAGGAAGATCTTTGGGGATTGTAAAGAGGAAG  
AACTGTTGCCCAAAGCTATGGATGAATCTGTTGCAAGAAAACCTGGGATATCAGTGAAGTG  
ATGGTTGGCCTGCTAAAATAGGAAACAAGGAGTAAAGAGCTGTTTATAAACTGCATATCAG  
TTATATCTGTGATCAGGAATGGTGTGGATTGAGAACTTGTTACTTGAAGAAAAAGAATTTTG  
ATATTGGAATAGCCTGCTAAGAGGTACATGTGGGTATTTTGGAGTTACTGAAAAATATTTT  
TGGGATAAGAGAATTTTTCAGCAAAGATGTTTAAATATATATAGTAAGTATAATGAATAATAA  
GTACAATGAAAAATACAATTATATTGTAATAATTATACTGGGCAAGCATGGATGACATATTA  
ATATTGTGAGAATTAAAGTACTCAAAGTGCTATCGAGAGGTTTTTCAAGTATCTTTGAGTT  
TCATGGCCAAAGTGTTAACTAGTTTTACTACAATGTTTGGTGTGTTGTGTGGAATATCTGCG  
CTGGTGTGTGCACACAAGTCTTACTTGAATAAATTTACTGGTAC

## **FIGURE 122**

</usr/seqdb2/sst/DNA/Dnaseqs.min/ss.DNA58747

<subunit 1 of 1, 336 aa, 1 stop

<MW: 36865, pI: 9.15, NX(S/T): 2

MAVATAAAVLAAALGGALWLAARRFVGPRVQRLRRGGDPGLMHGKTVLITGANSGLGRATAAE  
LLRLGARVIMGCDRARAEEAAGQLRRELQAACGPEPGVSGVGELIVRELDLASLRSVRA  
FCQEMLQEEPRLDVLINNAGIFQCPYMKTEDGFEMQFGVNHGLGHFLLTNLLGLLKSSAPSR  
IVVSSKLYKYGDINFDDLNSEQSYNKSFCYSRSLANILFTRELARRLEGNTVTVNVLHPG  
IVRTNLGRHIHIPLLVKPLFNLVSWAFFKTPVEGAQTSIYLASSPEVEGVSGRYFGDCKEEE  
LLPKAMDESVARKLWDISEVMVGLLK

**Important features:**

**Signal peptide:**

amino acids 1-21

**Short-chain alcohol dehydrogenase family protein**

amino acids 134-144, 44-56 and 239-248

**N-glycosylation site.**

amino acids 212-215 and 239-242

## **FIGURE 123**

GGGGATTGTAAAGAGGAAGNACTGTGCCCAAAGNTATGGATGAATCTGTTGCAAGAAAATTN  
TGGGATATCAGTGAAGTGATGGTTNGCCTGCTAAAAATAGGAACAAGGAGTAAAAGAGCTGTT  
TATAAACTGCATATCAGTTATATCTGTGATCAGGAATGGTGTGGATTGAGAACTTGTTACT  
TGAAGAAAAAGAATTTTGATATTGGAATAGCCTGNTAAGAGGNACATGTGGGTATTTTGAG  
TTACTGAAAAATTATTTTGGGATAAGAGAATTTTCAGCAAAGATGTTTTAAATATATATAGT  
AAGTATAATGAATAATAAGTACAATGAAAAATACAATTATATTGTAAAATTATAACTGGGCA  
AGCATGGATGACATATTAATATTTGTGAGAATTAAGTGACTCAAAGTGCTATCGAGAGGTTT  
TTCAAGTATCTTTGAGTTTCATGGCCAAAGGTGTTAACTAGTTTTACTACAATGTTTGGTGTT  
TGTGTGGAATTATCTGCCTGGCTT

10017081-102400

## FIGURE 124

GAGAGGACGAGGTGCCGCTGCCTGGAGAATCCTCCGCTGCCGTCCGCTCCCGAGCCCAGCC  
CTTTCCTAACCCAACCCAACCTAGCCAGTCCCAGCCGCCAGCGCTGTCCCTGTACGGAC  
CCCAGCGTTACCATGTCATCTGCCGCTTCTCTATCCTTACCCGACCTCAGATGCTCCCTTCT  
GCTCCTGGTAACTTGGGTTTTTACTCCTGTAACTGAAATAACAAGTCTTGCTACAGAGA  
ATATAGATGAAATTTTAAACAATGCTGATGTTGCTTTAGTAAATTTTATGCTGACTGGTGT  
CGTTTCAGTCAGATGTTGCATCCAATTTTTGAGGAAGCTTCCGATGTCAATAAGGAAGAATT  
TCCAAATGAAATCAAGTAGTGTGGCCAGAGTTGATTGTGATCAGCACTCTGACATAGCCC  
AGAGATACAGGATAAGCAAATACCCAACCTCAAATGTTTTCGTAATGGGATGATGATGAAG  
AGAGAATACAGGGGTGAGCATGAGTAAAGCATTGGCAGATTACATCAGGCAACAAAAAG  
TGACCCCATTCAGAAATTCGGGACTTAGCAGAAATCACCCTCTTGATCGCAGCAAAAGAA  
ATATCATTGGATATTTGAGCAAAAGGACTCGGACAACATAGAGTTTTTGAACGAGTAGCG  
AATATTTTGCATGATGACTGTGCCCTTTCTTTCTGCATTTGGGGATGTTTCAAACCGGAAAG  
ATATAGTGGCGACAACATAATCTACAAACCACGAGGCATTCTGCTCCGATATGGTGTACT  
TGGGAGCTATGACAAATTTTGATGTGACTTACAATTGGATTCAAGATAAATGTGTTCTCTCT  
GTCCGAGAAATAACATTTGAAAAATGGAGAGGAATTGACAGAAGAAGGACTGCCTTTTCTCAT  
ACTCTTTCACATGAAAGAAGATACAGAAAGTTTAGAAATATCCAGAATGAAGTAGCTCGGC  
AATTAATAAGTGAAAAAGGTACAATAAACTTTTTACATGCCGATTGTGACAAATTTAGACAT  
CCTCTTCTGCACATACAGAAAACCTCCAGCAGATTGTCCGTGTAATCGCTATTGACAGCTTTAG  
GCATATGTATGTGTTTGGAGACTTCAAAGATGTATTAATTCCTGGAAAACCAAGCAATTCG  
TATTTGACTTACATTCGAAAACTGCACAGAGAATTCATCATGGACCTGACCCAACTGAT  
ACAGCCCCAGGAGAGCAAGCCCAAGATGTAGCAAGCAGTCCACCTGAGAGCTCCTTCCAGAA  
ACTAGCACCCAGTGAATATAGGTATACTCTATTGAGGGATCGAGATGAGCTTTAAAAAATTG  
AAAAACAGTTTGTAAGCCTTTCAACAGCAGCATCAACCTACGTGGTGGAAATAGTAAACCTA  
TATTTTCATAATCTATGTGTATTTTTATTTTGAATAAACAGAAAGAAATTTAAAAA  
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

1007687-102401

## FIGURE 125

</usr/seqdb2/sst/DNA/Dnaseqs.min/ss.DNA57689  
<subunit 1 of 1, 406 aa, 1 stop  
<MW: 46927, pI: 5.21, NX(S/T): 0  
MHPAVFLSLPDLRCSLLLLVTWVFTPVTTEITSLATENIDEILNNADVALVNFYADWCRFSQ  
MLHPIFEEASDVKEEFNENQVVFARVDCDQHS DIAQRYRISKYPTLKLFRNGMMMKREYR  
GQRSVKALADYIRQQKSDPIQEIRDLAEITTLDRSKRNIIGYFEQKDSDNRYRVFERVANILH  
DDCAFLSAFGDVSKPERYSGDNIIYKPPGHSAPDMVYLGAMTNFVDVTYNWIQDKCVPLVREI  
TFENCEELTEEGLPFLILFHMKEDTESLEIFQNEVARQLISEKGTINFLHADCDKFRHPLLH  
IQKTPADCPVIAIDSPRHMYVFGDFKDVLPGLKQFVFDLHSGKLRHREFHHGPDPTDTAPG  
EQAQDVASSPPESFQKLAPSEYRYTLRDRDEL

### **Important features:**

#### **Signal peptide:**

amino acids 1-29

#### **Endoplasmic reticulum targeting sequence.**

amino acids 403-406

#### **Tyrosine kinase phosphorylation site.**

amino acids 203-211

#### **Thioredoxin family proteins**

amino acids 50-66

## FIGURE 126

ATTAGGAAGAATTTCCAAATGAAAATCAAGTAGTNTTTGCCAGAGTNGATTGTGATCAGCA  
CTCTGACATAGCCCAGAGATACAGGATAAGCAAATACCCAACCCCTCAAATTGTTTCGTAATG  
GGATGATGATGAAGAGAGAATACAGGGGTGAGCGATCAGTGAAAGCATTGGCAGATTA

10017081-102401

## FIGURE 127

AGAGGCCTCTCTGGAAGTTGTCCCGGGTGTTGCGCCGCGGAGCCCGGGTCGAGAGGACNAGG  
TGCCGCTGCTGGAGAATCCTCCGCTGCGCTCGGCTCCCGGAGCCAGCCCTTTCCTAACCC  
AACCCAACTAGCCCGTCCCGAGCGCCAGCGCCTGTCCCTGTGTCGCGANCCAGCGTNACC  
ATGCATCCTGCCGTCTTCTCTATCCTTACCCGACCTCAGATGCTCCCTTCTGCTCCTGGTAAC  
TTGGGTTTTTACTCCTGTAACAACGAAATAACNNGTCTTGATACNNAGAATATAGATGAAA  
TTTTAAACNATGCTGATGTGGCTTTAGTCAATTTTTATGCTGACTGGTGTGCTTCAGTCAG  
ATGTGGCATCCAATTTTTGAGGANGCTTCCGATGTCATTAAGGAAGAATTTCCAAATGAAAA  
TCAAGTAGTGTGTTGCCAGAGTTGATTGTGATCAGCACTCTGACATAGCCAGAGATACAGGA  
TAAGCAAATACCCAACCCTCAAATTGTTTCGTAATGGGATGATGATGAAGAGAGAATACAGG  
GGTCAGCGATCAGTGAAAGCATTGGCAGATTACATCAGGC

1007091-102401

## FIGURE 128

GCCACGCGTCCGATGGCGTTACGTTGCGGCGCTTCTGCTACATGCTGGCGCTGCTGCTCA  
CTGCCGCGCTCATCTTCTTCGCCATTTGGCACATTATAGCATTGATGAGCTGAAGACTGAT  
TACAAGAAATCCTATAGACCAGTGTAAATACCCCTGAATCCCCCTTGTACTCCAGAGTACCTCAT  
CCACGCTTTCTTCTGTGTCATGTTTCTTTGTGCAGCAGAGTGGCTTACACTGGGTCTCAATA  
TGCCCCCTCTTGGCATATCATATTTGGAGGTATATGAGTAGACCAGTGTAGTGGCCAGGA  
CTCTATGACCCCTACAACCATCATGAATGCAGATATTTCTAGCATATTGTGAGAAGGAAGGATG  
GTGCAAATTAGCTTTTTATCTTCTAGCATTTTTTTACTACCTATATGGCATGATCTATGTTT  
TGGTGAGCTCTTAGAACAACACAGAAGAATTGGTCCAGTTAAGTGCATGCAAAAAGCCAC  
CAAAATGAAGGGATTCTATCCAGCAAGATCCTGTCCAAGAGTAGCCTGTGGAATCTGATCAGT  
TACTTTAAAAAATGACTCCTTATTTTTTAAATGTTTCCACATTTTGTCTGTGAAAGACTG  
TTTTTCATATGTTTATACTCAGATAAAGATTTTAAATGGTATTACGTATAAATTAATATAAAAT  
GATTACCTCTGGTGTTGACAGGTTTGAACCTTGCACITCTTAAGGAACAGCCATAATCCTCTG  
AATGATGCATTAATTACTGACTGTCCTAGTACATTGGAAGCTTTTGTTTTAGGAACTTGTA  
GGGCTCATTTTGGTTTCATTGAAACAGTATCTAATTATAAATTAGCTGTAGATATCAGGTGC  
TTCTGATGAAGTGAAAATGTATATCTGACTAGTGGGAACTTCATGGGTTTCCCTCATCTGTC  
ATGTCGATGATTATATATGGATACATTTACAAAAATAAAAAGCGGGAATTTTCCCTTCGCTT  
GAATATTATCCCTGTATATTGCATGAATGAGAGATTTCCCATATTTCCATCAGAGTAATAAA  
TATACTTGCTTTAATTCTTAAGCATAAGTAAACATGATATAAAAATATATGCTGAATTACTT  
GTGAAGAATGCATTTAAAGCTATTTTAAATGTGTTTTTATTTGTAAGACATTACTTATTAAG  
AAATTGTTATTATGCTTACTGTTCTAATCTGGTGGTAAGGTATTCTTAAGAATTGTCAGG  
TACTACAGATTTTCAAACCTGAATGAGAGAAAATTGTATAACCATCCTGCTGTTCTTTAGT  
GCAATACAATAAACTCTGAAATTAAGACTC

## FIGURE 129

</usr/seqdb2/sst/DNA/Dnaseqs.min/ss.DNA23330  
<subunit 1 of 1, 144 aa, 1 stop  
<MW: 16699, pI: 5.60, NX(S/T): 0  
MAFTFAAFCYMLALLLTAALIFFAIWHIIAFDELKTDYKNPIDQCNTLNPLVLPEYLIHAFF  
CVMFLCAAEWLTLGLNMPLLAYHIWRYMSRPVMSGPLYDPTTIMNADILAYCQKEGWCKLA  
FYLLAFFYYLYGMIYVLVSS

### **Important features:**

#### **Signal peptide:**

amino acids 1-20

#### **Type II transmembrane domain:**

amino acids 11-31

#### **Other transmembrane domain:**

amino acids 57-77 and 123-143

## FIGURE 130

ATTATAGCATTGATGAGCTGAAGACTGATTACAAGATCCTATAGACCAGTGAATACCCTG  
AATCCCCTTGTACTCCCAGAGTACCTCATCCACGCTTCTTCTGTGTCATGTTCTTTGTGC  
AGCAGAGTGGCTTACACTGGGTCTCAATATGCCCCCTCTTGGCATATCATATTTGGAGGTATA  
TGAGTAGACCAGTGATGAGTGGCCAGGACTCTATGACCCTACAACCATCATGAATGCAGAT  
ATTCTAGCATATTGTCAGAAGGAAGGATGGTGCAAATTAGCTTTTATCTTCTAGCATTTTT  
TTACTACCTATATGGCATGATCTATGTTTTGGTGAGCTCTTAGAACACACACAGAAGAATT  
GGTCCAGTTAAGTG CATGCAAAAAGCCACCAAATGAAGGGATTCTATCCAGCAAGATCCTGT  
CCAAGAGTAGCCTGTGGAATCTGATCAGTTACTTTAAAAAATG

FIGURE 130

## FIGURE 131

CGGACGCGTGGGGGAAACCTTCCGAGAAAACAGCAACAAGCTGAGCTGCTGTGACAGAGGG  
GAACAAGATGGCGGCGCCGAAGGGGAGCCTCTGGGTGAGGACCCAACCTGGGGCTCCCGCCG  
TGCTGCTGCTGACCATGGCCTTGGCCGGAGGTTTCGGGGACCGCTTCGGCTGAAGCATTTGAC  
TCGGTCTTGGGTGATACGGCGTCTTGCCACCGGGCCTGTCAGTTGACCTACCCCTTGACAC  
CTACCCTAAGGAAGAGAGTTGTACGCATGTCAGAGAGTTGCAGGCTGTTTTCAATTTGTC  
AGTTTGTGGATGATGGAATTGACTTAAATCGAATAAATTGGAATGTGAATCTGCATGTACA  
GAAGCATATTTCCCAATCTGATGAGCAATATGCTTGCCATCTTGGTTGCCAGAATCAGCTGCC  
ATTGCTGAACTGAGACAAGAACAACCTTATGTCCCTGATGCCAAAATGCACCTACTCTTTC  
CTCTAACTCTGGTGAGGTCACTTCTGGAGTGACATGATGGACTCCGCACAGAGCTTCATAACC  
TCTTCATGGACTTTTTATCTTCAAGCCGATGACGGAAAAATAGTTATATTCCAGTCTAAGCC  
AGAAATCCAGTACGCACCACATTTGGAGCAGGAGCCTACAAATTTGAGAGAATCATCTCTAA  
GCAAAATGTCCTATCTGCAAATGAGAAATTCACAAGCGCACAGGAATTTTCTTGAAGATGGA  
GAAAGTGATGGCTTTTTAAGATGCCTCTCTCTTAACTCTGGGTGGATTTTAACTACAACCTCT  
TGTCTCTCGGTGATGGTATTGCTTTGGATTGTTGTGCAACTGTTGCTACAGCTGTGGAGC  
AGTATGTTCCCTCTGAGAAGCTGAGTATCTATGGTGACTTGGAGTTTATGAATGAACAAAAG  
CTAAACAGATATCCAGCTTCTTCTTGTGGTGTGTAGATCTAAAACCTGAAGATCATGAAGA  
AGCAGGGCCTCTACCTACAAAAGTGAATCTTGCTCATTCTGAAATTTAAGCATTTTTCTTTT  
AAAAGACAAGTGAATAGACATCTAAAATTCACCTCCTCATAGAGCTTTTAAAAATGGTTTCA  
TTGGATATAGGCCTTAAGAAATCACTATAAAATGCAATAAAGTTACTCAAATCTGTG

10017081.102441

## FIGURE 132

</usr/seqdb2/sst/DNA/Dnaseqs.min/ss.DNA26847

<subunit 1 of 1, 323 aa, 1 stop

<MW: 36223, pI: 5.06, NX(S/T): 1

MAAPKGSLSWVRTQLGLPPLLLLLTMALAGGSGTASAEAFDSVLGDTASCHRAQLTYPLHTYP  
KEEELYACQRCRLFSICQFVDDGIDLNRKLECESACTEAYSQSDEQYACHLGCQNQLPFA  
ELRQEQLMSLMPKMHLLFPLTLVRSFWSMDMSAQSFITSSWTFYLQADDGKIVIFQSKPEI  
QYAPHLEQEPTNLRESSLSKMSYLMRNSQAHRNFLEDGESDGFRLCLSLNSGWILTTTLVL  
SVMVLLWICCATVATAVEQYVPSEKLSIYGDLEFMNEQKLNRYPASSLVVVRSKTEDHEEAG  
PLPTKVNLAHSEI

**Important features:**

**Signal peptide:**

amino acids 1-31

**Transmembrane domain:**

amino acids 241-260

**N-glycosylation site.**

amino acids 90-93

10017031.103401

TTGGGTGATACGGCGTCTTGCCACCGGGCTGTGAGTTGACCTACCCCTTGACACCTACCC  
TAAGGAAGAGGAGTTGTACGATGTGAGAGGTTGCAGGCTGTTTTCAATTTGTGAGTTG  
TGGATGATGGAATTGACTTAAATCGAACTAAATTGGAATGTGAATCTGCATGTACAGAAGCA  
TATTTCCCAATCTGATGAGCAATATGCTTGCCATCTTGTTGGCCAGAATCAGCTGCCATTGCG  
TGAACTGAAGCAAGAACAACATATGTCTCTGATGCCAAAATGCACCTACTCTTTCTCTAA  
CTCTGGTGAGGTCATTCTGGATGACATGATGAACCTCCG

## **FIGURE 134**

CACACTGGCCGGATCTTTTAGAGTCCTTTGACCTTGACCAAGGGTCNGGAAAACAGCAACAA  
GCTGAGCTGCTGTGACAGAGGGGAACAAGATGGCGGCGCCGAAGGGAGCCTTTGGGTGAGGAC  
CCAACTGGGGCTCCC GCCGTGCTGCTGCTGACCATGGCCTTGCCCGAGGTTTCGGGGACCG  
CTTCGGCTGAAGCATTGACTCGGTCTTGGGTGATACGGCGTCTTGCCACCGGGCCTGTCAG  
TTGACCTACCCCTTGCAACCTACCCTAAGGAAGAGGAGTTGTACGCATGTCAGAGAGGTTG  
CAGGCTGTTTTCAATTTGTCAGTTTGTGGATGATGGAATTGACTTAAATCGAACTAAATTGG  
AATGTGAATCTGCATGTACAGAAGCATATTCCCAATCTGATGAGCAATATGCTTGCCATCTT  
GGTTGCCAGAATCAGCTGCCATTGCTGAAGTGAAGACAAGAACAACCTTATGTCCCTGATGCC  
AAAAATGCACCTACTCTTTCCTCTAACTCTGGTGAGGTCATTCTGGAGTGACATGATGGACT  
CCGC

CCGC

## FIGURE 135

GCAGAGTGGCGATCGCTGAGAGGCAGGAGGGCCGAGGCGGGCCTGGGAGGCGGCCCGGAGGT  
GGGGCGCCGCTGGGGCCGGCCCCGCACGGGCTTCATCTGAGGGCGCACGGCCCGCGACCGAGC  
GTGCGGACTGGCCCTCCCAAGCGTGGGGCGACAAGCTGCCGGAGCTGCAATGGGCGCGCGCTG  
GGGATTCCTTTTGGCCCTCCTGGGCGCCGTGTGGCTGCTCAGCTCGGGCCACGGAGAGGAGC  
AGCCCCCGGAGACAGCGGCACAGAGGTGCTTCTGCCAGGTTAGTGTTACTTGGATGATTGT  
ACCTGTGATGTTGAAACCATTGATAGATTTAATAACTACAGGCTTTTCCCAAGACTACAAA  
ACTTCTTGAAAGTGACTACTTTTAGGTATTACAAGGTAAACCTGAAGAGCCCGTGTCTTTCT  
GGAATGACATCAGCCAGTGTGGAAGAAGGCACTGTGCTGTCAAACCATGTCAATCTGATGAA  
GTTCTGATGGAATTAATCTGCGAGCTACAAGTATTCTGAAGAAGCCAATAATCTCATTGA  
AGAATGTGAACAAGCTGAACGACTTGGAGCAGTGGATGAATCTCTGAGTGAGGAAACACAGA  
AGGCTGTTCTTCAGTGGACCAAGCATGATGATTCTTCAGATAACTTCTGTGAAGCTGATGAC  
ATTGAGTCCCTGAAGCTGAATATGTAGATTTGCTTCTTAATCCTGAGCGCTACACTGGTTA  
CAAGGGACCAGATGCTTGGAAAATATGGAATGTCACTACGAAGAAAAGTGTTTTAAAGCCAC  
AGACAATTAAGAAGACCTTTAAATCCTTTGGCTTCTGGTCAAGGGACAAGTGAAGAGAACACT  
TTTTACAGTTGGCTAGAAAGTCTCTGTGTAGAAAAAGAGCATTCTACAGACTTATATCTGG  
CCTACATGCAAGCATTAAATGTGCATTTGAGTGAAGATATCTTTTACAAGAGACCTGGTTAG  
AAAAGAAATGGGGACACAACATTACAGAAATTTCAACAGCGATTTGATGGAATTTTGACTGAA  
GGAGAAGGTCGAAGAAGGCTTAAGAACTGTATTTTCTCTACTTAATAGAACAAGGGCTTT  
ATCCAAGTGTTACCATTTCTCGAGCGCCAGATTTTCAACTCTTTACTGGAATAAAATTC  
AGGATGAGGAAAAAATAATGTACTTCTGGAAATACCTCATGAATCAAGTCAATTTCTTTG  
CATTTTGATGAGAATTCATTTTTTGTCTGGGATAAAAAAGAGCACAAAATAAAGGAGGA  
CTTTCGACTGCATTTTAGAAATATTTCAAGAATTATGGATTGTGTTGGTTGTTTTAAATGTC  
GTCTGTGGGGAAGCTTCAGACTCAGGGTTTGGGCACTGCTCTGAAGATCTTATTTTCTGAG  
AAATGTATAGCAAATATGCCAGAAAGTGGACCTAGTTATGAATTCATCTAACCAGACAAGA  
AATAGTATCATTATTCACAGCATTGGAAGAATTTCTACAAGTGTGAAAGAATTAGAAAAT  
TCAGCAATCTGTTACAGAATATTCATTAAAGAAAACAAGCTGATATGTGCCTGTTTCTGGAC  
AATGGAGGCGAAAGAGTGGAAATTTCAATCAAAGGCATAATAGCAATGACAGTCTTAAGCCAA  
ACATTTTATATAAAGTTGCTTTTGTAAAGGAGAATTATATGTTTTAAGTAACACATTTTT  
AAAAATTGTGTTAAGTCTATGTATAATACTACTGTGAGTAAAGTAATCTTTAATAATGTG  
GTACAAATTTTAAAGTTAATATTGAATAAAAGGAGGATTATCAAATTAATAAAAAAAAAA  
AAAAAAAAAAAAAAAAAAAAAAAAA

## **FIGURE 136**

</usr/seqdb2/sst/DNA/Dnaseqs.min/ss.DNA53974  
<subunit 1 of 1, 468 aa, 1 stop  
<MW: 54393, pI: 5.63, NX(S/T): 2  
MGRGWGFLFGLLGAVWLLSSGHGEEQPPETAARCFQCQVSGYLDDCTCDVETIDRFNNYRLF  
PRLQKILLESDFRYKVNLRKPCPFWNDISQCGRRDCAVKPCQSDEVDPGIKSASYKYSEEA  
NNLIEECEQAERLGAVDESLSEETQKAVLQWTKHDDSSDNFCEADDIQSPEAEYVDLLLNPE  
RYTGYKGPDAWKIWNVIYEENCFKPQTIKRPLNPLASGQGTSEENTFYSWLEGLCVEKRAFY  
RLISGLHASINVHLSARYLLQETWLEKKWGHNITEFQQRFDGILTEGEGPRRLKNLYFLYLI  
ELRALSQVLPFFFERPDFQLFTGNKIQDEENKMLLLEILHEIKSFPLHFDENSFFAGDKKEAH  
KLKEDFRLHFRNISRIMDCVGCFCRLWGKLQTQGLGTALKILFSEKLIANMPESGSPSYEFH  
LTRQEIIVSLFNAFGRISTSVKELENFRNLLQNIH

### **Important features:**

#### **Signal peptide:**

amino acids 1-23

#### **N-glycosylation site.**

amino acids 280-283 and 384-387

#### **Amidation site.**

amino acids 94-97

#### **Glycosaminoglycan attachment site.**

amino acids 20-23 and 223-226

#### **Aminotransferases class-V pyridoxal-phosphate**

amino acids 216-222

#### **Interleukin-7 proteins**

amino acids 338-343

## FIGURE 137

GCTGGAAATATGGATGTCATCTACGAGAAACTGTTTTAAGCCACAGACAATTTAAAGACCTT  
TAAATCCTTTGGCTTCTGGTCAAGGGACAAGTGAAGAGNACACTTTTTACAGTTGGCTAGAA  
GGTCTCTGTGTAGAAAAAGAGCATTTCTACAGACTTATATCTGGCCTACATGCAAGCATTAA  
TGTGCATTTGAGTGCAAGATATCTTTTACAAGAGACCTGGTTAGAAAAGAAATGGGGACACA  
ACATTACAGAATTTNAACAGCGATTTGATGGAATTTTGACTGAAGGAGAAGGTCCAAGAAGG  
CTTAAGAACTTGATTTTTCTCTACTTAATAGAACTAAGGGCTTTATCCAAAGTGTTACCAT  
CTTNGAGCGCCAGATTTTCAACTNNTTACTGGAAATAAAATTCAGGATGAGGNAAAACAAA  
TGTTACTTTTGGAAATACTTCATGAAATCAAGTCATTTTCCTTTGCATTTTGATGAGAATTCA  
TTTTTTTGCTG

104201-102401

## FIGURE 138

CGGACGCGTGGGCGGACGCGTGGGCGGACGCGTGGGTTGGGAGGGGGCAGGATGGGAGGGAA  
AGTGAAGAAAAACAGAAAAGGAGAGGGACAGAGGCCAGAGGACTTCTCATACTGGACAGAAAC  
CGATCAGGCATGGAACTCCCTTCGTCTACTCACCTGTTCTTGCCCTGGTGTTCCTGACAGG  
TCTCTGCTCCCTTTTAACTGGATGAACATCACCCACGCCTATTCCAGGGCCACCAGAAG  
CTGAATTTGGATACAGTGTCTTACAACATGTTGGGGGTGGACAGCGATGGATGCTGGTGGGC  
GCCCCCTGGGATGGGCCTTCAGGCGACCGAGGGGGGACGTTTATCGCTGCCCTGTAGGGGG  
GGCCCACAATGCCCATGTGCCAAGGGCCACTTAGGTGACTACCAACTGGGAAATTCATCTC  
ATCTGTCTGTGAATATGCACCTGGGGATGTCTCTGTTAGAGACAGATGGTGATGGGGGATTC  
ATGGTGAGCTAAGGAGAGGGTGGTGGCAGTGTCTCTGAAGGTCCATAAAAGAAAAAGAGAA  
GTGTGGTAAGGGAAAATGGTCTGTGTGGAGGGGTCAAGGAGTTAAAAACCTAGAAAGCAAA  
AGGTAGGTAATGTCAGGGAGTAGTCTTCATGCCTCCTTCAACTGGGAGCATGTTCTGAGGGT  
GCCCTCCAAGCCTGGGAGTAACATATTTCCCCCATCCCCAGGCCTGTGCCCTCTCTGGTCT  
CGTGCTTGTGGCAGCTCTGTCTTCAGTTCTGGGATATGTGCCCGTGTGGATGCTTCATTCCA  
GCCTCAGGGAAGCCTGGCACCCACTGCCCAACGTGAGCCAGAGGAAGGCTGAGTACTTGGTT  
CCCAGAAGGAGATACTGGGTGGGAAAAAGATGGGGCAAAGCGGTATGATGCCTGGCAAAGGG  
CCTGCATGGCTATCCTCATTGCTACCTAATGTGCTTGCAAAAGCTCCATGTTTCTTAACAGA  
TTCAGACTCCTGGCCAGGTGTGGTGGCCACACCTGTAATTCTAGCACTTTGGGAGGCCAAG  
GTGGGCAGATCACTTGAGGTGAGGAGTTCAAGACCAGCCTGGCCAACATGGTGAAACTCCAT  
CTCTACTAAAAAATAACAAAAATTAGCTGGGTGCGCTAGTGCATGCCTGTAATCTC  
ATCTACTCGGGAGGCTAAGACAGGAGACTCTCACTTCAACCCAGGAGGTGGAGGTTGCGGTG  
AGCCAAGATTGTGCCCTCTGCACCTAGCGTGGGTGACAGAGTAAGCGAGACTCCATCTCAA  
AATAATAATAATAATAATTCAGACTCCTTATCAGGAGTCCATGATCTGGCCTGGCACAGTAA  
CTCATGCCTGTAATCCCAACATTTTGGGAGGCCAACGCAGGAGGATTGCTTGAGGTCTGGAG  
GTTTGAGACCAGCCTGGGCAACATAGAAAGACCCCATCTCTAAATAAATGTTTTAAAAAT

## **FIGURE 139**

```
></usr/seqdb2/sst/DNA/Dnaseqs.min/ss.DNA57039
><subunit 1 of 1, 124 aa, 1 stop
><MW: 13352, pI: 5.99, NX(S/T): 1
MELPFVTHLFLPLVFLTGLCSFNLDEHHPRLPFGPPEAEFGYSVLQHVGGGQRWMLVGAPW
DGPSGDRRGDVYRCFVGGAHNAPCAKGHLGDYQLGNSSHPAVNMHLGMSLLETGDGGFMVS
```

### **Important features:**

#### **Signal peptide:**

amino acids 1-22

#### **Cell attachment sequence.**

amino acids 70-73

#### **N-glycosylation site.**

amino acids 98-101

#### **Integrins alpha chain proteins**

amino acids 67-81

## **FIGURE 140**

CACAGTTCCTCCACCATCACTCCTCCATTCTTCCAACCTTTATTTTCTAGCTTGCCATTGGGA  
GGGGGCAGGATGGGAGGGAAAGTGAAGAAAACAGAAAAGGAGAGGGACAGAGGCCAGAGGAC  
TTCTCATACTGGACAGAAACCGATCAGGCATGGAACCTCCCTTCGTCACTCACCTGTTCTTG  
CCCTGGTGTTCTCTGACAGGTCTCTGCTCCCCCTTTAACCTGGATGAACATCACCACGCCT  
ATTCCCAGGGCCACCAGAAGCTGAATTTGGATACAGTGTCTTACAACATGTTGGGGGTGGAC  
AGCGATGGATGCTGGTGGGCGCCCCCTGGGATGGGCCTTCAGGCGACCGGAGGGGGGACGTT  
TATCGCTGCCCTGTAGGGGGGGCCCAATGCCCCATGTGCCAAGGGCCACTTAGGTGACTA  
CCAACTGGGAAATTCATCTCATCCTGCTGTGAATATGCACCTGGGGATGTTCTGTTAGAGA  
CAGATGGTGATGG

10017081-102403

## FIGURE 141

AAAGTTACATTTTCTCTGGAACCTCTCCTAGGCCACTCCCTGCTGATGCAACATCTGGGTTTG  
GGCAGAAAGGAGGGTGCTTCGGAGCCCGCCCTTTCTGAGCTTCCCTGGGCCGGCTCTAGAACA  
ATTGAGGCTTCGCTGCGACTCAGACCTCAGCTCCAACATATGCATTCTGAAGAAAAGATGGCT  
GAGATGGACAGAATGCTTTATTTTGGAAAGAAACAATGTTCTAGGTCAAACCTGAGTCTACCA  
AATGCAGACTTTCACAATGGTTCTAGAAGAAATCTGGACAAGTCTTTTCATGTGGTTTTTCT  
ACGCATTGATTCCATGTTTGCTCAAGATGAAGTGGCCATTCTGCCTGCCCTCAGAACCTC  
TCTGTACTCTCAACCAACATGAAGCATCTCTTGATGTGGAGCCAGTGATCGCGCCTGGAGA  
AACAGTGTACTATTCTGTGGAATACCAGGGGGAGTACGAGAGCCTGTACACGAGCCACATCT  
GGATCCCCAGCAGCTGGTGCTCACTCACTGAAGGTCTGAGTGTGATGTCACTGATGACATC  
ACGGCCACTGTGCCATACAACCTTCGTGTGAGGGCCACATTGGGCTCAGACCTCAGCCTG  
GAGCATCCTGAAGCATCCCTTTAATAGAACTCAACCATCCTTACCCGACCTGGGATGGAGA  
TCACCAAAGATGGCTTCCACCTGGTTATGTAGCTGGAGGACCTGGGGCCCCAGTTTGAGTTC  
CTTGTCGCCCTACTGGAGGAGGGAGCCTGGTGCCGAGGAACATGTCAAATGGTGAGGAGTGG  
GGGTATTCCAGTGCACTTAGAAACCATGGAGCCAGGGCTGCATACTGTGTGAAGGCCCAGA  
CATTCGTGAAGGCCATTGGGAGGTACAGCGCCTTCAGCCAGACAGAATGTGTGAGAGTGC  
GGAGAGGCCATTCCCTGGTACTGGCCCTGTTTGCCCTTTGTTGGCTTCATGCTGACTCTGT  
GGTCGTGCCACTGTTCTGCTGGAATAATGGCCGGCTGCTCCAGTACTCCTGTTGCCCGTGG  
TGCTCCTCCAGACACCTTGAAAAAACAACATTCACCCAGAAAGTTAATCAGCTGCAGAAGG  
GAGGAGGTGGATGCCTGTGCCACGGCTGTGATGTCTCCTGAGGAACTCCTCAGGGCTGGAT  
CTCATAGGTTTGCGGAAGGGCCAGGTGAAGCCGAGAACCTGGTCTGCATGACATGGAACC  
ATGAGGGGACAAGTTGTGTTTTCTGTTTTCCGCCACGGAACAAGGGATGAGAGAAGTAGGAAGA  
GCCTGTTGTCTACAAGTCTAGAAGCAACCATCAGAGGCAGGGTGGTTTGTCTAACAGAAAC  
TGACTGAGGCTTAGGGGATGTGACCTCTAGACTGGGGGCTGCCACTTGCTGGCTGAGCAACC  
CTGGGAAAAGTGACTTCATCCCTTCGGTCTCAAGTTTCTCATCTGTAATGGGGGAATTACC  
TACACACCTGCTAAACACACACACAGAGTCTCTCTATATATACACACGTACACATAAA  
TACACCCAGCACTTGCAAGGCTAGAGGGAACTGGTGACACTCTACAGTCTGACTGATTAG  
TGTTTCTGGAGAGCAGGACATAAATGTATGATGAGAATGATCAAGGACTCTACACACTGGGT  
GGCTTGGAGAGCCCACTTTCCAGAAATATCCTTGAGAGAAAAGGAATCATGGGAGCAATGG  
TGTTGAGTTCATTCTCAAGCCCAATGCCGTGCAGAGGGGAATGGCTTAGCGAGCTCTACAGT  
AGGTGACCTGGAGGAAGGTACAGCCCACTGAAAATGGGATGTGCATGAACCGGAGGATC  
CATGAACACTGTAAAGTGTGACAGTGTGTGCACTGCAGACAGAGGTGAAATGTATGT  
GTGCAATGCGACGAGAATGCAGAAGTCAGTAACATGTGCATGTTTGTGTGCTCCTTTTTC  
TGTTGGTAAAGTACAGAATTCAGCAAATAAAAAGGGCCACCCTGGCCAAAAGCGGTAAAAA  
AAAAAAAAA

1007/051.102401

## **FIGURE 142**

</usr/seqdb2/sst/DNA/Dnaseqs.min/ss.DNA57033

<subunit 1 of 1, 311 aa, 1 stop

<MW: 35076, pI: 5.04, NX(S/T): 2

MQTFMTVLEEIIWTS LFMWFFYALIPCLLTDEVAILPAPQNL SVLSTNMKHL LMWSPVIAPGE  
TVYYSVEYQGEYESLYTSHIWIPSSWCSL TEGPECDVTDITATVPYNLRVRATLGSQTS AW  
SILKHPFN RNSTILTRPGMEITKDG FHLVIELEDLGPQFEFLVAYWRREP GAEHHV KMVRSG  
GIPVHLETMEPGAAYCVKAQTFVK AIGRYSAFSQTECVEVQGEA IPLVLALFAFVGFM LILV  
VVPLFVWKMGRL LQYSCCPVVVLPDTLKITNSPQKLISCRREVDACATAVMSPEELLRAWIS

**Important features:**

**Signal peptide:**

amino acids 1-29

**Transmembrane domain:**

amino acids 230-255

**N-glycosylation site.**

amino acids 40-43 and 134-137

**Tissue factor proteins.**

amino acids 92-119

**Integrins alpha chain proteins**

amino acids 232-262

1661/081-102401

## FIGURE 143

TCCTGCTGATGCACATCTGGGTTTGGCAAAAGGAGGTTGCTTCGAGCCGCCCTTTCTAGCTT  
CCTGGCCGGCTCTAGAACAATTGAGGCTTCGCTGCGACTAGACCTCAGCTCCAACATATGCA  
TTCTGAAGAAAGATGGCTGAGATGACAGAATGCTTTATTTTGAAAGAAACAATGTTCTAGG  
TCAAACCTGAGTCTACCAAATGAGACTTTCACAATGGTTCTAGAAGAAATCTGGACAAGTCT  
TTTCATGTGGTTTTTCTACGCATTGATTCCATGTTTGCTCACAGATGAAGTGGCCATTCTGC  
CTGCCCCCTCAGAACCTCTCTGTACTCTCAACCAACATGAAGCATCTCTTGATGTGGAGCCCA  
GTGATCGCGCCTGGAGAAACAGTGTACTATTCTGTGGAATACAGGGGGAGTACGAGAGCCT  
GTACACGAGCCACATCTGGATCCCCAGCAGCTGGTGCTCACTCACTGAAGGTCCTGAGTGTG  
ATGTCACCTGATGACATCACGGCCACTGTGCCATACAACCTTTGTGTCAGGGCCACATTGGGC  
TCACAGACCTCAGCCTGGAGCATCCTGAAGCATCCCTTTAATAGAACTCAACCATCCTTAC  
CCGACCTGGGATGGAGATCACCAAAGATGGCTTNCACCTGGTTATTGAGCTGGAGGACCTGG  
GGCCCCAGTTTGAGTTTCCTTGTGGCCTANTGGAGGAGGGCGAACCCCTTGCGGCGCAAGGG  
GTTNGCGAACCCTTGCGGCCGCTGGGGTATCTCTCGAGAAAAGAGAGGCCCAATATGACCC  
ACATACTCAATATGGACGAANTGCTATTGTCCACCTGTTTGAGTGGCGCTGGGTTGAT

## FIGURE 144

CCCACGCGTCCGCCACGCGTCCGAGGGACAAGAGAGAAGAGAGACTGAAACAGGGAGAAGA  
GGCAGGAGAGGAGGAGGTGGGGAGAGCACGAAGCTGGAGGCCGACACTGAGGGAGGGCGGGA  
GGAGGTGAAGAAGGAGAGAGGGGAGAAGAGGCAGGAGCTGGAAAGGAGAGAGGGAGGAGGAG  
GAGGAGATGCGGGATGGAGACCTGGAGTTAGGTGGCTTGGGAGAGCTTAATGAAAAGAGAAC  
GGAGAGGAGGTGTGGGTTAGGAACCAAGAGGTAGCCCTGTGGGCAGCAGAAGGCTGAGAGGA  
GTAGGAAGATCAGGAGCTAGAGGGAGACTGGAGGGTTCGGGAAAAGAGCAGAGGAAAGAGG  
AAAGACACAGAGAGACGGGAGAGAGAAGAAGAGTGGGTTTGAAGGGCGGATCTCAGTCCCTG  
GCTGCTTTGGCATTGGGGAACTGGGACTCCCTGTGGGGAGGAGAGGAAAGCTGGAAAGTCCCT  
GGAGGACAGGGTCCAGAAGGAGGGGACAGAGGAGCTGAGAGAGGGGGGAGGGCGTTGGG  
CAGGGGTCCCTCGGAGGCCCTCTGGGGATGGGGGCTGCAGCTCGTCTGAGCGCCCTCGAGC  
GCTGTACTCTGGGCTGACTGGGGGCAGCAGCTCACATCGGACCAGCACTGACCCCGAGG  
ACTGGTGGAGCTACAAGGATAATCTCCAGGGAACTTCGTGCCAGGGCCTCCTTTCTGGGGC  
CTGTTGAATGCAGCGTGGAGTCTGTGTGCTGTGGGGAAGCGGCAGAGCCCCGTGGATGTGGA  
GCTGAAGAGGGTTCCTTTATGACCCCTTTCTGCCCCATTAAAGGCTCAGCACTGGAGGAGAGA  
AGCTCCGGGGAACCTTGTAACAACCCGGCCGACATGTCTCCTTCTGCTGCACCCCGACCT  
GTGGTCAATGTGTCTGGAGGTCCCTCCTTTACAGCCACCGACTCAGTGAAGTGGGGTGCT  
GTTTGGAGCTCGCGACGGAGCCGGCTCGGAACATCAGATCAACCACCGGGCTTCTCTGCTG  
AGGTGCAGCTCATTCACTTCAACCAGGAACTCTACGGGAATTTAGCGCTGCTCTCCCGCGG  
CCCAATGGCCTGGCCATTCTCAGCCTCTTTGTCAACGTTGCCAGTACCTCTAACCCATTCT  
CAGTCGCCTCCTTAACCGGCACACCATCACTCGCATCTCTACAAGAATGATGCCTACTTTC  
TTCAAGACCTGAGCCTGGAGCTCCTGTTCCCTGAATCCTTCGGCTTCATCACTATCAGGGC  
TCTCTCAGACCCCGCCCTGCTCCGAGACTGTACCTGGATCCTCATTGACCGGGCCCTCAA  
TATCACTCCCTTCAGATGCATCCCTGAGACTCCTGAGCCAGAATCCTCCATCTCAGATCT  
TCCAGAGCCTCAGCGGTAAACAGCCGGCCCTGCGAGCCCTTGGCCCACAGGGCACTGAGGGGC  
AACAGGGACCCCCGGCACCCCGAGAGCGCTGCGGAGGCCCAACTACCGCTGCATGTGGA  
TGGTGTCCCCCATGGTCGCTGAGACTCCCTTCGAGGATTGACCCGCGCCTCTAAGCCTC  
CCCACAAGGCGAGGGGAGTTACCCCTAAACAAGCTATTAAAGGGACAGAATACTTA

1007631-103401

## FIGURE 145

</usr/seqdb2/sst/DNA/Dnaseqs.min/ss.DNA34353  
<subunit 1 of 1, 328 aa, 1 stop  
<MW: 36238, pI: 9.90, NX(S/T): 3  
MGAAARLSAPRALVLWAAALGAAAHIGPAPDPEDWWSYKDNLQGNFVPGPPFWGLVNAAWSLC  
AVGKRQSPVDVELKRVLYDPFLPPLRLSTGGEKLRGTLYNTGRHVSFLPAPRPVVNVSGGPL  
LYSHRLSELRLFLGARDGAGSEHQINHQGFSAEVQLIHFNQELYGNFSAASRGPNGLAILSL  
FVNASTSNPFLSRLLNRDITITRISYKNDAYFLQDLSLELLFPESFGFITYQGSLSSTPPCSE  
TVTWTILIDRALNITSLQMHSLRLLSQNPSSQIFQSLSGNSRPLQPLAHRALRGNRDPRHPER  
RCRGPNYRLHVDGVPHGR

### **Important features:**

#### **Signal peptide:**

amino acids 1-23

#### **Transmembrane domain:**

amino acids 177-199

#### **N-glycosylation site.**

amino acids 118-121, 170-173 and 260-263

#### **Eukaryotic-type carbonic anhydrases proteins**

amino acids 222-270, 128-164 and 45-92

# FIGURE 146

GGCGCTGTTCTCGCGCTACTGGCTGTACGGAGCAGGAGCAAGAGGTGCGCGCCAGCCTCCGCCGCCGAGCCTC  
 GTTCGTGTCCTCCCGCCTCTGCTCTGTCAGCTACTGCTCAGAAACGCTGGGGCGCCACCTCTGGCAGACTAACGRA  
 GCAGCTCCCTTCCCACTCCCACTCGAGGTCTAATTTTGGACGCTTTGCTTGCCATTTCTTCCAGTTTGGAGGAGC  
 CGCAGAGGCGGAGGCTCGGATATCTCTGTCAGTCAGCACCCACGTCGCCCCCGAGCGCTCGTGCTCAGCGCCTTC  
 GCGAGCGGGGCTCTCCGTCTGCGGTCCCTTTGTGAAGGCTCTGGCGCGCTGCAGAGGCCGCGCCGCTCCGGTTTGGCT  
 CACCTCTCCAGGAACTTTCACACTGGAGAGCGAAAGGAGTGGAAAGAGCTGTCTGGAGATTTTCTCTGGGGA  
 ATCCTCAGGTCACTTATTTAGAGTGTACCGCGCGGGAGTGGCTCAGAGTAACCAACAGCTGCTTTATGGCTAGA  
 GCAATTCAGCCATGCTGGTTCCTCAATGCCATTTTATTGGAGAACTTTTGGAAAAATACATGGATGAGGATGCT  
 GAGTGGTGATAGCCAAACAAAGAGGGGAAAGGGCCATCAGAGCAATGACATGCAAGATTTTGGACCTTTCAT  
 AATAAATACGAAGTCAAGTGTATCCAAAGCCTCTAATATGGAGTATATGACATGGGATGTAGAGCTGGAAAGA  
 TCTCGAATCTCTGGGCTGAAAGTTGCTTGGGAAACATGGACCTGCAAGCTTGCTTCCATCAATTTGGACAGAT  
 TTTGGAGCACACTGGGGAAGATATAGGCCCGCGACGTTTCTATGTAACAATCGTGGTATGATGAAGTGAAAGACTTT  
 AGCTACCCATATGAACATGAATGCAACCCATATTGTCCATTGAGGTGTTCTGGCCCTGTATGTACACATTATACA  
 CAGGTCTGTGGGCACTAGTAACAGAATCGGTTGTGCCATTAAATTTGTGTCATAACATGAACATCTGGGGCGAG  
 ATATGGCCCAAGCTGTCTACCTGGTGTGCAATTTACTCCCAAAGGGGAACTGGTGGGGCCATGCCCTTTACAA  
 CTAGGGCGGCCCTGTTCTGCTTGCCCACTAGTTTGGAGGGGGCTGTAGAGAAAAATCTGTGCTACAAAGAGGG  
 TCAGACAGGTATTTCCCTCTGAGAGAGGAAACAAATGAATAGAACGACAGCAGTCACAAGTCCATGACACC  
 CATGTCCGCAAGATCAGATGATAGTAGCAGAAATGAAGTCATAGCGCACAGCAAAATGTCCCAAATTTGTTCT  
 TGTGAAGTAAATTAAGAGATCAGTGCAAGGAAACAACTGCAATAGGTACGAAATGTCTGCTGGCTGTGGGAT  
 AGTAAGCTAAAGTATTGGCAGTGTAATTATGAATGCAATCCAGCATCTGTAGAGCTGCAATTCATTATGATG  
 ATAAATAGACCAATGATGCTGGTGGGTAGATATCAGTACAGGAAAGAAAGCATTATTTTCATCAAGTCCAATAGA  
 AATGGTATTCAAACAAATTTGCCAAATATCAGTCTGCTAATTCCTTCAAGCTCTTAAAGTCAACAGTTTCAGGCTGTG  
 ACTGTGAAACCACTGTGGAAACAGCTCTGTCCATTTTCATAAGCCTGCTTCAAGTCCCAAGGTATACCTGTCTCT  
 CGTAACGTATATGCAAGCAATCCACATTATGCTCGTGTAAATTTGGAACCTCAGATTTTATCTGATCTGTGCAATATC  
 TGCAGAGCAGCATACATGCTCGAGTGTGTTGAAATCAAGGTGTTATGTTGATGTAATGCTGTGGACAAAGA  
 AAGACCTCATTGCTCTTTTCAGATTTGGAATCTTCTCAGAAAGTTTACAGAACTCTCCAGGAGAAAGGCATT  
 AGAGTGTGTTGCTGTTGTGTTAACTGAAATCTTGGAGAGGACCAATAAGCACTATTCCAAATGCCAATTTCTG  
 ATTTTGTATAAACTGTAACATTTACTGTACAGAGTACATCAACTATTTTCAGCCCAAAAGGTGCCAAATGCATA  
 TAAATCTTGATAAAACAAAGTCTATAAAATAAAACATGGGACATTAGCTTTGGGAAAAAGTAAATGAAAATATAATG  
 TTTTGAAGATCTCTGTGTTAAATATTGCTATAATTTCTTAGCAGTTATTTCACAGTTAATTACATAGTCATGATT  
 GTTCTACGTTTCATATATATATGTTGCTTTGTATATGCCACTAATAAAATGAATCTAAACATGAATGTGAATG  
 GCCCTCAGAAAAATCATCTAGTGTCTTAAATAATATCGACTCTAAAACTGAAAGAACTTATCACATTTTCCCC  
 AGTTCAATGCTATGCCATTACCAACTCCAAATAATCTCAATAAATTTTCCACTTAATAACTGTAAGGTTTTTTTC  
 TGTTAATTTAGGCGATATAGAATATTAAATCTGATATTGCACITCTTATTTTATATAAAATAATCTTTTAATATC  
 CAAATGAATCTGTGTTAAATGTTTGAATCTCTGGGAATGGCCTTAAAAATAAATGTAAATAAGTCAGAGTGGTGGT  
 ATGAAAACTTCTTATGATCATGTAGTAAATGTAGGGTTAAGCATGGACGACGAGCTTTCTATGTACTGTTA  
 AAATTGAGGTACATATTTCTTTTGTATCCTGGCAAAATCTCCTGCGAGCCAGGAAGTATAGTAGCAAAAGTT  
 GAAACAAAGATGAACATTAATGTATTACATTACCATTGGCACTGATTTTTTTTAAATGGTAAATGACCTTTGATATAA  
 ATATTGCGCATATCATGTACTATAATGGTGATATAATTTGTTCTATGAAAAATGTATTGTGCTTTGATACTAA  
 AATCTGTAATAATGTTAGTTTGGTAAATTTTTTTCTGCTGTGGATTACATATAAAATTTTTTCTGCTGTGGTGA  
 TAAACATTAAATTAATCATGTTTCAAAAAA

10017681.162401

## **FIGURE 147**

</usr/seqdb2/sst/DNA/Dnaseqs.min/ss.DNA45417

<subunit 1 of 1, 500 aa, 1 stop

<MW: 56888, pI: 8.53, NX(S/T): 2

MKCTARELRLVTTVLFMARAIPAMVVPNATLLEKLLLEKYMDEEDGEWWIAKQRGKRAITDNDM  
QSILDLHNKLRSQVYPTASNMEYMTWDVELERSAESWAESCLWEHGPA SLLPSIGQNLGAHW  
GRYRPPTTFHVQSWYDEVKDFSYPYEHECNPYCPFRCSGPVCTHYTQVVWATSNRIGCAINLC  
HNMNIWGQIWPKAVYLVLCNYSKGNWWGHAPYKHGRPCSACPPSFGGGCRENL CYKEGSDRY  
YPPREETNEIERQQSQVVDTHVRTRSDSSRNEVIS AQQMSQIVSCEVRLRDQCKGTT CNR  
YEC PAGCLDSKAKVIGSVHYEQSSICRAAIHYGIIDNDGGWVDITRQGRKH YFIKSNRNGI  
QTIGKYQSANSFTVSKVTQAVTCETTVEQLCPFHKPASHCPRVYCP RNCMQANPHYARVIG  
TRVYDLSSICRAAVHAGVVRNHGGYVDVMPVDKRKTYIASFQNGIFSESLQNPPGGKAPRV  
FAVV

**Important features:**

**Signal peptide:**

amino acids 1-20

**Extracellular proteins SCP/Tpx-1/Ag5/PR-1/Sc7 protein**

amino acids 165-186, 196-218, 134-146, 96-108 and 58-77

**N-glycosylation site**

amino acids 28-31

## FIGURE 148

GCGGAGACAAGCGCAGAGCGCAGCGCACGGCCACAGACAGCCCTGGGCATCCACCGACGGCG  
CAGCCGGAGCCAGCAGAGCCGGAAGGCGCGCCCCGGGCAGAGAAAGCCGAGCAGAGCTGGGT  
GGCGTCTCCGGGCGCCGCTCCGACGGGCCAGCGCCCTCCCCATGCTCCCTGCTCCACGCCG  
GCCCCCTCCGGTCAGCATGAGGCTCCTGGCGGCCGCTGCTCCTGCTGCTGCTGGCGCTGT  
ACACCGCGCGTGTGGACGGGTCCAAATGCAAGTGCTCCCGGAAGGGACCCAGATCCGCTAC  
AGCGACGTGAAGAAGCTGGAATGAAGCCAAAGTACCCGCACTGCGAGGAGAAGATGGTTAT  
CATCACCACCAAGAGCGTGTCCAGGTACGAGGTGAGGAGCACTGCCTGCACCCCCAAGCTGC  
AGAGCACCAAGCGCTTCATCAAGTGGTACAACGCCCTGGAAACGAGAAGCGCAGGGTCTACGAA  
GAATAGGGTGAAAAACCTCAGAAGGGAAAACTCCAAACCAAGTTGGGAGACTTGTGCAAAGGA  
CTTTGCAGATTAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGCCTTTC  
TTTCTCACAGGCATAAGACACAAATTATATATGTTATGAAGCACTTTTACCAACGGTCAG  
TTTTTACATTTTATAGCTGCGTGCAGAAAGGCTTCAGATGGGAGACCCATCTCTCTGTGCT  
CCAGACTTCATCACAGGCTGCTTTTTATCAAAAAGGGGAAAACTCATGCCCTTTCCTTTTAA  
AAAATGCTTTTTTGTATTTGTCCATACGTCACTATACATCTGAGCTTTATAAGCGCCCGGA  
GGAACAATGAGCTTGGTGGACACATTTTCATTGCAGTGTGCTCCATTCTAGCTTGGGAAGC  
TTCGCTTAGAGGTCTGGCGCCTCGGCACAGCTGCCACGGGCTCTCCTGGGCTTATGGCCG  
GTCACAGCCTCAGTGTGACTCCACAGTGGCCCCCTGTAGCCGGGCAAGCAGGAGCAGGTCTCT  
CTGCATCTGTTCTCTGAGGAACTCAAGTTTGGTTGCCAGAAAAATGTGCTTCATCCCCCCT  
GGTTAATTTTTACACACCCTAGGAAACATTTCCAAGATCCTGTGATGGCGAGACAAATGATC  
CTTAAAGAAGGTGTGGGGTCTTCCCAACCTGAGGATTTCTGAAAGGTTACAGGTTCAATA  
TTTAAATGCTTCAGAAGCATGTGAGGTTCCCAACACTGTGAGCAAAAACCTTAGGAGAAAACT  
TAAAAATATATGAATACATGCGCAATACACAGCTACAGACACACATTCTGTTGACAAGGAA  
AACCTTCAAAGCATGTTTCTTCCCTCACCACAACAGAACATGCAGTACTAAAGCAATATAT  
TTGTGATTTCCCATGTAAATCTTCAATGTTAAACAGTGCAGTCTCTTTTCGAAAGCTAAGAT  
GACCATGCGCCCTTTCCTCTGTACATATACCTTAAAGAACGCCCTCCACACACTGCCCC  
CAGTATATGCCGCACTGTACTGCTGTGTTATATGCTATGTACATGTGAGAAACATTAGCAT  
TGCATGCAGGTTTCATATTTCTTCTAAGATGGAAAGTAATAAAATATATTTTGAAATGTAAAA  
AAAAAAAAAA

## **FIGURE 149**

MSLLPRRAPPVSMRLAALLLLLLALYTARVDGSKCKCSRKGPKIRYSDVKKLEMKPKYPH  
CEEKMWIITTKSVSRYRGQEHCLHPKLQSTKRFIKWYNANNEKRRVYEE

**Signal sequence:**

amino acids 1-34

# FIGURE 150

GCCCAGGGAGCTGCTATGGCTTCCTTTGTTGTTCACCCCGGTCTGCGCTCATGTTAAACTCCAATGTCTCCTGTG  
 GTTAACCTGCTCTTGGCCATCAAGTTCAACCCTCATGACAGCCAGCAGTATCCAGTTGTCAACCAAAATATGAG  
 CAAAATCCGGGGCTAAGAAACCGTTACCCAAATGAGATCTTGGGTCCAGTGGAGCAGATCTTAGGGGTCCCCTA  
 TGCTTCACCCCCACTGGGAGAGAGGGCGGTTTCAGCCCCCAGAACCCCCGTCTCTCTGAGCTGGCATCCGAATATC  
 TACTCAGTTTGTCTGTCTGTGTGCCCCAGCAGCTGGATGAGAGATCCTTACTGCATGCATGTCTGCCCCATCTGTTT  
 TACCGCCAATTTGGATACCTTTTGATGACCTATGTTCAAGATCAAAATGAAGACTGCTTCTTACTTAAACATCTACGT  
 GCGCCAGAGATGGAGCCCAACAAGAAAAACGAGATGATATAACGAGTAATGACCGTGGTGAAGACGAAGA  
 TATTATGATCAGAACAGTAAGAAGCCCGTTCATGGTCTATATCCATGGGGGATCTTACATGGAGGGGACCGGCAA  
 CATGATTGACGGCAGCATTTTGGCAAGCTACGAAACGTCATCGTGATCACCATTAACCTACCGTCTGGGAATACT  
 AGGGTTTAAAGTACCGGTGACAGGCAGCAAAAGCAACTATGGGCTCCTGGATCAGATTCAAGCACTGCGGTG  
 GATTGAGGAGATGCGGAGCCTTTGGCGGGGACCCCAAGAGATGACCATCTTTGGCTCGGGGGCTGGGGCCTC  
 CTGTGTGAGCCTGTTGACCCCTGTCCCACTACTCAGAAAGTCTCTTCCAGAAGGCCATCATTAGAGCGGGCACCGC  
 CCTGTCCAGCTGGGCAGTGAATACAGCCGCGCAAGTACACTCGGATATTGGCAGACAAGGTGCGCTGCAACAT  
 GCTGGACACACGGAATGCTAGATGCTGCGGAAACAGAACTACAAGGAGCTCATCCAGCAGACCATCAACCC  
 GGCCACTACCACATAGCCTTGGGGCCGCTGATCGACGGCGACGTATCCACAGACGCCCCAGATCCTGATGGA  
 GCAAGGCGAGTTTCTCACTACGACATCATGCTGGGCGTCAACCAAGGGGAAGGCGTGAAGTTCTGTGAGCGCAT  
 CGTGGATAACGAGGACGCTGTGACGCCCAACGACTTTGACTTCTCGTGTCCAACCTTCGTGGACAACCTTTACGG  
 CTACCTCTGAAGGGAAAGACACTTTGCGGGAGACTATCAAGTTTCATGTACACAGACTGGGCGGATAAGGAAAAACC  
 GGAGACGCGCGGGAACCCCTGGTGGCTCTCTTTACTGACCACAGTGGGTGGCCCCCGCGTGGCCGCGGACT  
 GCACGCGCAGTACGGCTCCCCACCTACTTCTATGCTCTTATCATCACTGCAAGGCGAAATGAAGCCAGCT  
 GGCAGATTGCGGCCATGGTGATGAGGTCCCTATGCTTTCGGCATCCCATGATCGGTGCCACAGCTCTTCAG  
 TTGTAACCTTTTCAAGAACGACGTCTATGCTCAGCGCGTGGTTCATGACCTACTGGAGCAACTTCGCCAAAACTGG  
 TGATCCAAATCAACAGTTCTCAGGATACCAAGTTTCATTACACAAAACCCAGTCTTGAAGAAGTGGCGCTG  
 GTCCAAGTATATCTCCAAAGCCAGCTCTATCTGCATATTGGCTTGAACCCAGAGATGAGAGATCACTACCTGGC  
 AACGAAGTGTGCTTCCGTTTGAAGTCTGTTCTCTATTGGCACAACTTGAAGAGATTTCAAGTATGTTTTCAC  
 AACCAAAAGGTTCTCCACAGACATGACATCATTTCCCTATGGCACCCGCGATCTCCGCCAAGATATGCGC  
 AACCCAAAGCCGCAACCAATCACTCTGCGCAACATCCCAAACTACTAAGAACCCCTCAACCAAGGCCCTGA  
 GACACAACTGCTCTCATGAAACCAACAGAGATTATTCCACCGAATTAAGTGTCAACATTGCCCCGCGGCTG  
 GCTCCTCTTCTCTCAACATCTAGCTTTTGGGCGCTGTACTACAAAAGGACAAGAGGCGCCATGAGACTCACAG  
 GCGCCCCAGTCCCGAGAGAACACCAAAATGATATCGCTCACATCCAGAACGAAGAGATCATGTCCTTCGAGAT  
 GAAGCAGCTGGAACACGATCACAGGTGTGAGTCTGCTGACGGCACAACGACACTGAGGCTCACCTGCCCCGAGA  
 CTCACCCCTACGCTGCGCGGTGCGCAGATGACATCCCATTTATGACGCGCAACACCATCACCATGATTCCAAA  
 CACACTGACGSGATGACGCTTTGACACATTTTAAACCTTCAGTGGAGGACAAAACAGTACAAATTTACCCCA  
 CAGCATTTCCACCATAGATAGAGCTTTGCCCCATTTCCTTCTCTATCCCTCTGCCCCACCGCTCAGCAACAT  
 AGAAGAGGGAAGGAAAGAGAGAGGAAAGAGAGAGAGAGAAAGAGTCTCCAGACCGAAGATGTTTTGTGCCACT  
 GACTTAAGACAACAAATGCAAAAGGCGAGTCACTCCATCCCGGAGACCTTATCGTTGGTGTTTTCCAGTATTAC  
 AAGATCAACTTCTGACCTGTGAAATGTGAGAAGTACATTTCTGTTAAATAAAGTCTTAAAGATCTCTACCA  
 CTCCAATCAATGTTTAGTGATGATAGGACATCACCATTTCAAGGCCCGGGTGTTCACACGTCATGGAAGCAGCT  
 GACACTTCTGAACCTCAGCCAGGACACTTGATATTTTTTAATACAAATGGAAGTTTAAACATTTCTTCTGTG  
 CACACAATGGATGGCTCTCCTTAAGTGAAGAAAGAGTCAATGAGATTTTGCAGCAGCATGGAGCTGTAATCCAG  
 AGAAGAGGAAACGTAGAAATTTATTAATAAAGAAATGGAGCTGACAGCGAAATCTGTACGGTCTTGTGCAAGAG  
 GTGTTTTCGACGCTGAACTATTTAAGAGACTTTTG

40017081-102401

MLNSNVLLWLTLALAIKPTLIDSQAQYFVVNTNYGKIRGLRTPLPNEILGPVEQYLGVFPYASP  
PTGERRFPQPEPPSSWTGIRNTTQFAAVCPQHLDRSLLHMDLPWFWTANLDTLMTYVQDQN  
EDCLYLNIIYVPTEDGANTKKNADDITSNDRGEDEDIHDQNSKKPVMVYIHGSSYMEGTGNMI  
DGSILASYGNVIVITINYRLGILGFLSTGDQAAKGNYGLLDQIQALRWIEENVGAFGDPKR  
VTIFGSGAGASCVSLLTLTSHYSEGFLQKAI IQSGTALSSWAVNYQPAKYTRILADKVGCNML  
DTTDMVECLRNKNYKELIQQTITPATYHIAFGPVIDGVPIDDPQILMEQGEFLNYDIMLGV  
NQGEGLKFVDGIVDNEGDVTPNDFDFSVSNFVDNLYGYPEGKDTLRETIKFMYTDWADKENP  
ETRKRKTLVALFTDQHWPAPAAADLHAQYGSPTYFYAFYHHCQSEMPPSWADSAHGDEVFPYV  
FGIPMIGPTELFSCNFSKNDVMSLAVVMTYWTFNAKTGDPNPQVPQDTKFIHTKPNRFEeva  
WSKYNPKDQLYLHIGLKPRVRDRHYRATKVAFWLELVPHLHNLNEIFQYVSTTTKVPDPDMTS  
FPYGTRRSPAKIWPTTKRPAITPANPKHKSDDPKHTGPEDTTVLIETKRDYSTELSVTIAVG  
SLQLFLNLILAAALYYKDKRRHETHRRPSPQRNTTNDIAHIQNEEIMSLQMKQLEHDHECE  
ASLLEHDTLRLTCFPDYTLTLRRSPDDIPLMTPNTITMIPNTLTGMQPLHTFNTFSGGQNSTN  
LPHGHSSTRV

Signal sequence:

amino acids 1-24

Transmembrane domains:

amino acids 189-204, 675-692

## FIGURE 152

GGGAAAGATGGCGGGGACTCTGGGACCCCTTGGGTCGTGGCAGCAGTGGCGGCGATGTTTGT  
CGGCTCGGGATGGGTCCAGGATGTTACTCCTTCTTCTTTGTGGGGTCTGGGCAGGGGCCA  
CAGCAAGTCGGGGCGGGTCAAACGTTFCGAGTACTTGAACGGGAGCACTCGCTGTGCGAAGCC  
TACCAAGGGTGTGGGCACAGGCAGTTCCTCAGTGTGGAATCTGATGGGCAATGCCATGGTGA  
TGACCCAGTATATCCGCTTACCCAGATATGCAAGTAACAGGGTGCCCTGTGGAACCGG  
GTGCCATGTTTCTGAGAGACTGGGAGTTGCAGGTGCACCTTCAAATCCATGGACAAGGAAA  
GAAGAAATCTGCATGGGGATGGCTTGGCAATCTGGTACACAAGGATCGGATGCAGGCAGGGC  
CTGTGTTTGGAAACATGGACAATTTGTGGGGCTGGGAGTATTTGTAGACACCTACCCCAAT  
GAGGAGAAGCAGCAAGAGCGGGTATTCCCTACATCTCAGCCATGGTGAACAACGGCTCCCT  
CAGCTATGATCATGAGCGGGATGGGCGGCCATCAGAGCTGGGAGGCTGCACAGCCATTGTCC  
GCAATCTTCATTACGACACCTTCCTGGTGATTGCTACGTCACGCAAGAGGCATTTGACGATAATG  
ATGGATATTGATGGCAAGCATGAGTGGAGGGACTGCATTGAAGTGCCCGAGTCCGCTGCC  
CCGCGGCTACTACTTCGGCACCTCCTCCATCACTGGGGATCTCTCAGATAATCATGATGTCA  
TTTCTTGAAGTTGTTTGAAGTCGACAGTGGAGAGAACCCGGAAGAGGAAAAGCTCCATCGA  
GATGTGTTCTTGCCCTCAGTGGACAATATGAAGCTGCCTGAGATGACAGCTCCACTGCCGCC  
CTGAGTGGCCTGGCCCTTCTCCTCATCGTCTTTTCTCCTGTGTTTTCTGTATTTGGCA  
TAGTCATTGGTATCATACTCTACAACAAATGGCAGGAACAGAGCCGAAAGCGCTTCTACTGA  
GCCCTCCTGCTGCCACCCTTTTGTGACTGTCACCCATGAGGTATGGAAGGAGCAGGCACTG  
GCCTGAGCATGCAGCCTGGAGAGTGTCTGTCTCTAGCAGCTGGTTGGGGACTATATTCTG  
TCACTGGAGTTTTGAATGCAGGGACCCCGCATTCCTATGTTGTGTCATGGGGACATCTAATC  
CTGGCTCTGGGAAGCCACCACCCAGGGCAATGCTGCTGTGATGTGCCCTTCCCTGCACTCC  
TTCCATGTGGGAGCAGAGGTGTGAAGAGAATTTACGTGGTTGTATGCCAAAATCACAGAAC  
AGAATTTCATAGCCAGGCTGCCGTGTTGTTTGAAGTCTGAGAGGCCCTTCTACTTCAGTTTTG  
AATCCACAAAGAAATTAATACTGGTAACACCACAGGCTTTCTGACCATTCCATTCGTTGGGTT  
TGTCAATTTGACCAACCCCTCTGCCTACCTGAGGAGCTTTCTTTGGAAACCAGGATGGAACCT  
TCTTCCCTGCCTTACCTTCCCTTCACTCCATTCAATGTCTCTCTGTGTGCAACCTGAGCTG  
GGAAAGGCATTTGGATGCCTCTGTGGGGCTGGGGCTGCAGAACACACCTGCGTTTTCAC  
TGGCCTTCATTAGGTGGCCCTAGGGAGATGGCTTTCTGCTTTGGATCACTGTTCCCTAGCAT  
GGGTCTTGGGCTTATGGCATGTCCATGGCCTTCCCAATCAAGTCTCTTCAGGCCCTCAGTG  
AAGTTTGGCTAAAGGTTGGTGTAATAATCAAGAGAAGCCTGGAAGACATCATGGATGCCATG  
GATTAGCTGTGCAACTGACCAAGCTCCAGGTTTGATCAAACCAAAGCAACATTTGTCTATGTG  
GTCTGACCATGTGGAGATGTTTCTGGACTTGCTAGAGCCTGCTTAGCTAGCTGTTTTGTAGT  
TACGATTTTGAATCCCACTTTGAGTGTGAAAGTGAAGGAAGCTTTCTTCTTACACCTT  
GGGCTTGGATATTGCCAGAGAAGAAATTTGGCTTTTTTTTTCTTAATGGACAAGAGACAGT  
TGCTGTTCTCATGTTCCAAGTCTGAGAGCAACAGACCTCATCATCTGTGCTGGAGAGATT  
CACTGTCAATTGAGCAGCAGCCTGAGTGTGCGCTCTGTCAACCTTATTCTCACTGCGCTTA  
TTTGACAAGGGGTTACATGCTGCTGCTCACTTACTGCGCTGGGATTAACAGTTACAGGCCA  
AGTCTCCTTGGAGGGCTTGGAACTCTGAGTCTCCTATGAACCTCTGTAGCCTAAATGAAAT  
TCTTAAATACCCGATGGAAACCAAAAAAAAAAAAAAAAAAGGGCGCCGCACTCTAGAGTCG  
ACCTGCAGTAGGGATAACAGGGTAATAAGCTTGGCCGCATG

10017081-102401

## **FIGURE 153**

></usr/seqdb2/sst/DNA/Dnaseqs.min/ss.DNA50911  
><subunit 1 of 1, 348 aa, 1 stop  
><MW: 39711, pI: 8.70, NX(S/T): 1  
MAATLGPLGSWQQWRRCLSARDGSRMLLLLLLLGSGQGQPVGAGQTFEYLKREHSLSKPYQ  
GVGTGSSSLWNLMGNAMVMTQYIRLTPDMQSKQGALWNRVPCFLRDWELQVHFKIHGQGKKN  
LHGDGLAIWYTKDRMQPGPVFGNMDKFGVGLGVFVDTPNEEKQQQERVFPYISAMVNNGSLSY  
DHERDGRPTELGGCTAIVRNLHYDTFLVIRYVKRHLTIMMDIDGKHEWRDCIEVPGVRLPRG  
YYFGTSSITGDLSDNHDVISLKLFEFTVERTPEEEKLHRDVFLPSVDNMKLPENTAPLPPLS  
GLALFLIVFFSLVFSVFAIVIGIILYNKWQEQSRKRFY

**Signal sequence:**

amino acids 1-38

**Transmembrane domain:**

amino acids 310-329

# FIGURE 154

CCGAGCCGGGCGCGCAGCGACGGAGCTGGGCGCGCCTGGGACCATGGGCGTAGTGCAATCTACGGATCAGTCT  
 CTGATGCTGGGTCGTTAACCTCAGTGGGGAATCCAAAGATTCCATGAAGAAATCAGTTGTCTTCATTCAAGAAT  
 TGGGGTCTGGCTCAGAAATCTCTGCAGCTGGTGAAATCTGTTTCTAGAAGAGGTTTAATTAATGAGCTCAGGCTC  
 GACATGTTCCCGATTGGAGTGAACCATGAAGAGAAATAGAAATACTTAATATGCTTTCCGCAACCGCTTCT  
 TGCTGCTGCTGGCCCTGCGCTGCGCTGCGCTTTGTGAGCCTCAGCCTGCGAGTCTTCCACCTGATCTCCGGGTGT  
 CGACTCCTAAGAATGGAATGAGTAGCAAGAGTCAAGAGAAATCATGCCGAGCCCTGTGACGGAGCCCCCTGTGA  
 CAGACCCCGTTTATGAAGCTCTTTTGTACTGCAACATCCCAGTGTGGCCGAGCGCAGCATGGAAGGTGATGCC  
 CGCATCATTTAAGCTGCTCTCAGTGCAATGTTTTCATTCGCCACCGGAGACAGGTACCCACTGTATGTCTATCCCA  
 AAACAAGCGACCAAGAAATTGACTGCACTCTGCTGGCTAAACAGGAAACCGTATCACCCAAAGCTGGAAGCTTTCA  
 TTAGTGCATGTCAAAAGGATCCGAGCCCTCTTCGAAAGCCCTTGAACCTCTTGCCTCTTTACCCAAATCAAC  
 CATTGTGTGAGATGGGAGACTCACACAGAGGAGTTGTGCAAGCATTTGCAGAAACGGTGTGAGGCTGTGAGGATA  
 TCTATCTAAAGAAACAAAACTCCTGCCCAATGATGTGCTGCAGACAGCTCTATTTAGAGACCACTGGGAAAA  
 GCGCGACCTCAAAAGTGGGCTGGCCTTGTCTTTATGGCTTCTCCAGATTTTGACTGGAAGAAGATTTATTTCA  
 GGCACCAAGCAAGTGCCTGTTCTGCTCTGGAAGCTGCTATTGCCCGTAAGAAACAGTATCTGGAAGAGGAGC  
 AGCGTGTGAGTACCTCTACGTTTGAAGAAACAGCCAGCTGGAGAGACCTACGCGGAGATGGCCAAAGTCTGTGG  
 ATGTCCTCCACCAAGCAGCTTAGAGCTGCCAACCCCATAGACTCCATGCTCTGCCACTTCTGCCCAATGTGAGT  
 TTCCCTGTACCAAAATGGCTGTGTGACATGGAGCACTTCAAGGTAATTAAGACCCATCAGATCGAGGATGAAA  
 GGGAAAGACGGGAGAAGAAATTTGACTTCGGGTATTTCTCTCTGGGTGCCACCCCATCTGAAACCAACCATCG  
 GCCGGATGACGCGTCCACCGAGGGCAGGAAAGAAGAGCTCTTTGCCCTCTACTCTGCTCATGATGTCTACTGTG  
 CACCAGTTCTCAGTGCTTCGGGCTTTTCAGAAAGCAGGTTCCCAAGGTTTGACCGAGGTGATCTTTGAGCTTT  
 GGCAAGACAGAGAAAAGCCAGTGAACATTCGCTCGGATTTCTTCAATGGCGTGAATGCAATTCACACCTCT  
 TTTCTGCCAGACCCACCAAGGCTTCTCCCAAGCCCATGTGCCCGCTTGAAAACCTTGGTCCCTTTTGAAAA  
 GGGACATGTTTGTAGCCTCGGCTGGCAGTGGTACAAATTTATATGATGCATGTCAAGGAAAGGATTCTTGAAGG  
 TATGCAGTACAGCAGTATAGATTCATGCCAATACAGAGCATAGGGAAGGTCCACTTCTAGTTTTCGTGTATAC  
 TAAGGTTAGAAGATTTATGCTTTTAAAGGCTAAATATTGTTTGTGGGAACACAGATGGTGGGGTTGAACAGT  
 AAGCACATTCGTCGAATGTGGTACGTGAATTGCTTGGTACAAATGGCCAGTTCAAGAGGAATAGAGGTAATCTT  
 TATCATGCGCAGACTCTCGCTAGATGCCAGAAATATATAGTTCAAGACCTGAAAGTGGCAATCAAGTTTGCAC  
 TCTCTCGGCTGCCCCATGTACTATGTGATGGAACCAAGCACCTCAACCAAAATTTTAACTCTAGACATT  
 TTTACCTTGTCTTGTAAAGATTTCTTGAAGTGATTTATCTAAATAAAGGTTGGCAACCTTTTCTGTAAAGG  
 GCCAGTTGTAAATTTATCAGACTGTGGACCAAAAGGCCACATACAGTCTCTGTACATACTACTCAACTCTGT  
 TTTCTGAAGCAGGAAAGCCACACAGACAGTACATAAAGGAATATGTGTAGCTGGGTTCCACGGCCAGACAAACCA  
 GATGTGACCAAGACTTGGCCCTGGGCTGTAGTTTGTGACCCCTCATCTAAAAAATAGGCTATACATACAATTGC  
 ACTTCCAGCACTTTGAAGACAGTGTGAATACCAAGAAATTTTCAATGGTCTCCTCCAGTAACCTCTCTGTAGAAACA  
 CAGAAATTTGGTCTGTATCTGACACTAGAAACAAACCTTGAGGGTAAATAAACATTGAATTAAGATGAATCATAGAA  
 AACTGATTAGAAGAAATCTTGATGTTTATGATGATTTGTTGATGCAATTTAGTATTTTATAGTTTGAAGAAATATT  
 CTGCTGTAGTCTATTGCTGTATATGCTGAAATTTTGTATGCCATTTAGTATTTTATAGTTTGAAGAAATATT  
 TCTAAGACCAAGTTTATAGTACTCTTATTCCTGTAGTAATATTCAATTTGCTGTACCTGCTTGGTGGTTAGAAG  
 GAGGCTAGAAGATGAATTCAGGCACTTTCTTCCAATAAACTAATTAAGGCTCACTCCCTTTGACCAAGCTGTAGA  
 ACTGGATTCAATTTTAAACCAATTTTCATCAGTTTCAAAATGGTAAATTCGTAGTATTTTTAAATGCGTTTTGA  
 AGAATCTTGCATTAAGTATTTCAGACTTTTATAAGGTGTTTATATATTAGAAGCAATTAATAATTACATCTG  
 TGATTTCTGAACATAATGGTGCTAATTCAAGAAATGGAAGTGAAAGTGAGATTCTCTGTGTGCTCAGCATTC  
 AACTTTTCTCTGTTTGTGTTTGTGCTGTTGCAATTTGAATATGCTCTGTTCTATAAATAAATTTTGAAGATAA

10017031-102401

## **FIGURE 155**

></usr/seqdb2/sst/DNA/Dnaseqs.min/ss.DNA48329

><subunit 1 of 1, 480 aa, 1 stop

><MW: 55240, pI: 9.30, NX(S/T): 2

MLFRNRFLLLLLALAALLAFVSLSLQFFHLIPVSTPKNGMSSKSRKRIMPDVTEPPVTDPVY  
EALLYCNIPSPAERSMEGHAPHHFKLVSVHVFIRHGDRYPLYVIPTKRPEIDCTLVANRKP  
YHPKLEAFISHMSKSGSGASFESPLNSLPLYPNHPLCEMGELTQTGVVQHLQNGQLLRDIYLK  
KHKLLPNDWSADQLYLETTGKSRTLQSGLALLYGFLPDFDWKKIYFRHQPSALFCSGSCYCP  
VRNQYLEKEQRRQYLLRLKNSQLEKTYGEMAKIVDVPTKQLRAANPIDSMCHFCNVSFP  
TRNGCVDMEHFVKVIKTHQIEDERERREKKLYFGYSLGALHPIILNQTIGRMQRATEGRKEELF  
ALYSAHDVTLSPVLSALGLSEARFPFAARLIFELWQDREKPSEHSVRILYNGVDVTFHTSF  
CQDHHKRSKPKMCPLNLVRFVKRDMFVALGGSGTNYDACHREGF

**Signal sequence:**

amino acids 1-18

# FIGURE 156

AAAAAAGCTCACTAAAGTTTCTATTAGAGCGAATACGGTAGATTTCCATCCCCTTTTGAAGAAACAGTACTGTGGA  
 GCTATTTAAGAGATAAAAAACGAATATCCTTTCTGGGAGTTCAAGATTGGTGCAGTAATTTGGTTAGGACTCTGAGC  
 GCCGCTGTTTACCAATCGGGGAGAGAAAGCGGAGATCCTGCTCGCCTTGACAGCGCCTGAAGCACAAAGCAGAT  
 AGCTAGGAATGAACCATCTCCTGGGAGTATGTTGAAACACAGGAGGAGCTGACTTCCCACTGTGCCATTTCTAT  
 GGGCGAAGGAACCTGCTCCTGACTTCAGTGGTTAAGGGCAGAATTGAAAAATTAATTTCTGGAGGGAAGATAAGATTGAT  
 TCTTGGCGGACTGCAACCGGAGACTCAAAAGGGCTTGTCTCTGGGGAATCCTCCTGGGGACTCTGTGGGAGACCGG  
 ATGCAACCGAGATAGCGTATTACAGTTCCGGAAGAGCTGGAGAAAGGCTTAGGGTGGGGCAGACTCTCCAGGACCT  
 GGGGCTGGAGGCCCGGGGAGCTCGCGAGGCGGGAGTCCGCATCATCCCAGAGGTAGGACGAGCTTTTCCGCCCT  
 GAATCGCGAGCGGGCAGCTTGGTCAAGCGGGCAGGATAGACGGGAGGAGCTCTGTATGGGGGCCATCAAGTG  
 TCAATTAAATCTAGACATTTCTGATGGAGGATAAAGTGAAAAATATATGGAGTAGAAGTAGAAGTAAGGGACATTAA  
 CGACAATGCGCTTACTTTCTGTGAAAGTGAATTAGAAATAAAAAATTAGTGAATAATGACCGCACTGAGATGCGGTT  
 CCCTTACCCCAAGCCTGGGATCCGGATATCGGGGAAGAACTCTCTGCAGAGCTACGAGCTCAGCCGAACACTCA  
 CTTCTCCCTCATCGTGCAAAATGGAGCCGACGGTAGTAAGTACCCGGAATTGGTGCTGAAACGCGCCCTGGACCG  
 CGAAGAAAAGGCTGCTCACCACCTGCTCTACGGCTCCGACGGGGCGACCCGGTGCGCACAGGCACCGCGG  
 CATCCCGTGATGGTTCTGGATGCGGAACGACAAACGACACGCGTTTGTCTCAGCCGAGTACCGCGGAGCGTTCC  
 GGAGAACTTGCGCCTTGGGACGCGAGCTGCTTGTAGTCAAAGCTACCGACCTGACGAAGGAGTCAATGCGGAAGT  
 GAGGTATTCCTTCCGGTATGTGGACGACAGGCGGCCAAGTTTCAAAGTAGATTGTAATTGAGGGAACAATATC  
 AACAAATAGGGGAGTTGGACCAAGGAGTACAGATTCTACAGATGGAAGTGAAGCAATGGATAATGACAGGATA  
 TTCTGCGCGAGCCAAAGTCTCTGATCACTGTTCTGGAGTGAACGACAAATGCCCGAGAAGTGGTCTCACTCTCT  
 CGCCAGCTCGGTTCCCGAAAGCTCTCCAGAGGGACATTAATTGCGCTTTTAAATGTAAATGACCAAGATTCTGA  
 GGAACCGGACAGGTGATCTGTTTCATCCAAGGAAATCTGCCCTTTAAATTAGAAAAATCTTACGGAAATTAATA  
 TAGTTAGTCAAGACATAGTCTTGGATAGGGAACAGGTTCCCTAGTCAACAACTCAGCTGACCGCCACTCAGCG  
 GGGAAACCGCGCCCTATCCAAGGAACTCATATCTCGCTGAAGCTGGGAGACACCAAGCAACCGCGGCTCTT  
 CCCTCAGGCTCTTATCCGCTTATATCCAGAGAACAAATCCAGAGGAGTTTCCCTGCTCTGTGACCGCCCA  
 CGACCCGAGCTGTGAGAGAACGCGCAGATCACTTATTCCTGGCTGAGAACCAATCAACGGGCGAAGCTATC  
 GTCTACGTGTCCATCACTCCGACACTGGGCTACTGTATGCGCTGAGCTCCTTCCGCTACAGGAGTCTCCGAGA  
 CTTGCAAGTGAAAGTATGCGCGGGGACACCGGCAACCGCCCTCAGCAGCAAGCTGTGCTGAGCGCTGTCTGCT  
 GCTGGACGAAACGACATGCTGCGCGAGATCTCTGTACCCCGCCCTGCCCAAGAGCGTTCCTCTGGCGTGGAGCT  
 GGTCTCCCGCTCGCGAGGCGCGCTACTCTGCTGACCAAGTGTGGCGGTGACAGAGACTCCGCGCCAGAACGC  
 CATGCTGTCTTACCTGCTCTCAAGCGCAGCGAGCGGGAATCTTCTCGTGGGTCTGCACAGCGGCGAGGTGG  
 CAGCGCGAGCGCTGCTGACAGAGAGCGCTCAAGCAGAGCCTCGTAGTGGCGGTCCAGGCAACCGCGAGCC  
 CCCTCTCTCCGCACTGCTCAGCTCACCGTGGCGGCGACAGCATCCCGAAGTCTTGGCGGACTCTGGCAG  
 CTTGAGTCAAGCTGACTCTGAACCTCAGAGCTCACTCTGTACCTGGTGGTGGCGGTGGCGGCTCGGCTCTCTG  
 CGTCTTCTGGCTGCTGCTCAITCTGCTGCTGGCGCTCAGGCTCGCGGCTGGCACAAGTCAAGCTGCTGCGAGC  
 TTACAGAGCGGCTTGACAGGAGCGCGCGCTCGCACTTTTGGGCGTGGAGCGGGTCCAGGCTTTCTGCGAGAC  
 CTATTTCCCAAGAGTTTCCCTCACCACCGGACTCGCGGAAGAGTCACTGATTTCCCGCAGCCCAACTATGAGA  
 CATGCTCTGCTCAGCGAGGAGCTTTGAAAAAGCGAGCCCTTTTGTGTCAGGTGATTCGGTATTTTCTAAAGA  
 CAGTCTATGGTTAATTGAGGTGAGTTTATATCAAAATCTTCTTTCTTTTTTTTAAATGTCTGTCTCCCAAGC  
 TGGAGTGCAGCGGTACGATCATAGCTCACTGCGGCTCAAACTCCTAGGCTCAAGCAATATCCCACTTTGCTCT  
 CCGGTGTAAAGGGACTACAGGTGCAAGCACTACTGTCTGCTATCTATCTATCTATCTATCTATCTATCTAT  
 CTATCTCTATCTATCTATCTATCTTCTTGTACAGAGGGAGTCTCAGCGCTGTAATCCAGTACTTTGGGAGGC  
 CGAGGCGGTGGATCAACCTGAGGTTGGAGTTTGAGACCAAGCTGACCAACATGGAGAAACCCGCTTATACTAA  
 AAAAAATCAAAATAGCCGGGCGTGGTGGTGCATGCTCTGTAATCCAGCTACTTGGGAGGCTGAGTCAGGAGAAAT  
 TGCTTTAAGCTGGGAGGTGGAGGTTGCAATGAGCTGAGATTGTGCGATTGCTACTCCAGCTGGGCAACAAGAGTG  
 AACTCTATCTCA

10077001.102407

## **FIGURE 157**

>/usr/seqdb2/sst/DNA/Dnaseqs.min/ss.DNA48306  
>subunit 1 of 1, 916 aa, 1 stop  
>MW: 100204, pI: 4.92, NX(S/T): 4  
MIPARLHRDYKGLVLLGILLGTLWETGCTQIRYSVPPEEKEGSRVGDISRDLGLEPRELAER  
GVRIIPRGRTQLFALNPRSGSLVTAGRIDREELCMGAIKQLNLDILMEDKVKIYGVVEVR  
DINDNAPYFRESELEIKISENAATEMRFPPLPHAWDPDIGKNSLQSYELSPNTHFSLIVQNGA  
DGSKYPELVLKRALDREEKAAHHLVLTASDGGDPVRTGTARIRVMVLDANDNAPAFAPQPEYR  
ASVPENLALGTQLLVVNATDPDEGVNAEVRYSFYVDDKAAQVKLDCNSGTISTIGELDHE  
ESGFYQMEVQAMDNAGYSARAKVLITVLDVNDNAPEVVLTSLASSVPENSPRGTLIALLNVN  
DQDSEENGQVICFIQGNLPFKLEKSYGNYSYSLVTDIVLDREQVPSYINITVTATDRGTPPLST  
ETHISLNVADTNDNPPVFPQASYSAYIPENNPRGVSLSVTAHDPDCENAQITYSLAENTI  
QGASLSSYSVINS D TGVLYALSSFDYEQFRDLQVKVMARDNGHPLSSNVLSLFLVDQNDN  
APEILYPALPTDGSTGVELAPRSAEPGYLVTKVVAVD RDSQGNAWLSYRLLKASEPGLFSVG  
LHTGEVRTARALLDRDALKQSLVVAVQDHGQPPLSATVTTLTVAVADSIPQVLADLGSLESPA  
NSETSDLTLYLVAVA AVSCVFLAFVILL LALRLRRWHKSRL LQASGGGLTGAPASHFVGVD  
GVQAFLQTY SHEVSLTDSRKSHLIFPQPNYADMLVSQESFEKSEPLLLSGDSVFSKDSHGL  
IEVSLYQIFFLFFNC SVSQAGVQRYDHSSLRPQTPLKQLSHLCLRCNRDYRCKPPTVCLS  
IYLSIYLSIYLSIYLLSCTDGS LTPVIPVLWEAEAGGSPEVGSLRPA

### **Signal sequence:**

amino acids 1-30

### **Transmembrane domains:**

amino acids 693-711, 809-823, 869-888

## FIGURE 158

CCCAGGCTCTAGTGCAGGAGGAGAAGSAGGAGGAGCAGGAGGTGGAGATTCCAGTTAAAAG  
GCTCCAGAATCGTGTACCAGGCAGAGAACTGAAGTACTGGGGCCTCCTCCACTGGGTCCGAA  
TCAGTAGGTGACCCCGCCCTGGATTCTGGAAGACCTCACCATGGGACGCCCCGACCTCGT  
GCGGCCAAGACGTGGATGTTCTCTGCTCTTGCTGGGGGGAGCCTGGGCAGGACACTCCAGGGC  
ACAGGAGGACAAGGTGCTGGGGGGTCTAGTGCCAAACCCATTGCGAGCCTTGGCAGGCGG  
CCTTGTTCCAGGGCCAGCAACTACTCTGTGGCGGTGTCCTTGTAAGGTGGCAACTGGGTCTTT  
ACAGCTGCCCACTGTAAAAAACCGAAATACACAGTACGCCTGGGAGACCACAGCCTACAGAA  
TAAAGATGGCCAGAGCAAGAAATACCTGTGGTTCAGTCCATCCCACACCCCTGCTACAACA  
GCAGCGATGTGGAGGACCACAACCATGATCTGATGCTTCTTCAACTGCGTGACCAGGCATCC  
CTGGGGTCCAAAGTGAAGCCCATCAGCCTGGCAGATCATTGACCCAGCCTGGCCAGAAGTG  
CACCGTCTCAGGCTGGGGCACTGTCACCAGTCCCGAGAGAATTTCTGACACTCTCAACT  
GTGCAGAAGTAAAAATCTTTCCCCAGAAGAAGTGTGAGGATGCTTACCCGGGGCAGATCACA  
GATGGCATGGTCTGTGCAGGCAGCAGCAAAGGGGCTGACACGTGCCAGGGCGATTCTGGAGG  
CCCCCTGGTGTGTGATGGTGCACTCCAGGGCATCACATCCTGGGGCTCAGACCCCTGTGGGA  
GGTCCGACAAACCTGGCGTCTATACCAACATCTGCCGCTACCTGGACTGGATCAAGAAGATC  
ATAGGCAGCAAGGGCTGATATTCTAGGATAAGCACTAGATCTCCCTTAATAAACTCACAACCTCT  
CTGGTTC

## **FIGURE 159**

</usr/seqdb2/sst/DNA/Dnaseqs.min/ss.DNA48336  
<subunit 1 of 1, 260 aa, 1 stop  
<MW: 28048, pI: 7.87, NX(S/T): 1  
MGRPRPRAAKTWMFLLLLGGAWAGHSRAQEDKVLGGHECQPHSQPWQAALFQGQQLLCGGVL  
VGGNWWLTAAHCKKPKYTVRLGDHSLQNKDGPQEIPVVQSIHPFCYNSSDVEDHNNHDLMLL  
QLRDQASLGSKVKPISLADHCTQPGQKCTVSGWGTVTSPRENFPDTLNCAEVKIIFPQKKCED  
AYPGQITDGMVCAGSSKGADTCQGDGGPLVCDGALQGITSWGSDDPCGRSDKPGVYTNICRY  
LDWIKKIIGSKG

### **Important Features:**

#### **Signal peptide:**

amino acids 1-23

#### **Transmembrane domain:**

amino acids 51-71

#### **N-glycosylation site.**

amino acids 110-113

#### **Serine proteases, trypsin family, histidine active site.**

amino acids 69-74 and 207-217

#### **Tyrosine kinase phosphorylation site.**

amino acids 182-188

#### **Kringle domain proteins motif**

amino acids 205-217

# FIGURE 160

GCGCGCGGTGCACCGGGCGGGCTGAGCGCCTCCTGCGGGCCGGCCTGCGCGCCCCGGCCCCG  
 CGCGCCGCCACGCCCCAACCCCGGCCCGCGGCCCTTAGCCCCCGCCCGGGGCCCGCGCCCGC  
 GCCCGCGCCAGGTGAGCGCTCCGCCCGCGCGAGGCCCGCCCCGGGCCCGCCCCCGCCCCG  
 CCCCGCGCGCGGGGGAACCGGGCGGATTCTCGCGCGTCAAACCACTGATCCCATAAAAAC  
 ATTCATCTCTCCCGCGCGCCGCGCTGCGAGCGCCCCGCCAGTCCGCGCCCGCGCGCCCTCG  
 CCCTGTGCGCCCTGCGCGCCCTGCGCACCCGCGGCCGAGCCCCAGCCAGAGCCGGGCGGAGC  
 GGAGCGCGCCGAGCCTCGTCCCGCGGCCCGGGCGGGGCCGCTAGCGCGCGCGCTCGGA  
 TCGCGACCCGGCGCGGGGAGACGGGCGCCCGCCCCGAAACGACTTTCAGTCCCCGACGCGC  
 CCGGCCAACCCCTACGATGAAGAGGGCGTCCGCTGGAGGGAGCCGGCTGCTGGCATGGGTG  
 CTGTGGCTGCAGGCCTGGCAGGTGGCAGCCCCATGCCAGGTGCCTGCGTATGCTACAATGA  
 GCCCAAGGTGACGACAAGCTGCCCGCAGCAGGCCCTGCAGGCTGTGCCCGTGGGCATCCCTG  
 CTGCCAGCCAGCGCATCTTCTGCACGGCAACCGCATCTCGCATGTGCCAGCTGCCAGCTTC  
 CGTGCTGCGCAACCTCACCATCCTGTGGCTGCACTCGAATGTGCTGGCCCGAATTGATGC  
 GGCTGCCTTCACTGGCCTGGCCCTCCTGGAGCAGCTGGACCTCAGCGATAATGCACAGCTCC  
 GGTCTGTGGACCTGCCACATTCACGGCCTGGCGCGCTACACACGCTGCACCTGGACCGC  
 TCGCGCCTGCAGGAGCTGGGCCCGGGGCTGTTCCGCGGCCCTGGCTGCCCTGCAGTACCTCTA  
 CCTGCAGGACAACCGCGTGCAGGCACTGCCTGATGACACCTTCCGCGACCTGGGCAACCTCA  
 CACACCTCTTCTGCACGGCAACCGCATCTCCAGCGTGCCTGAGCGCGCCTTCCGTGGGGCTG  
 CACAGCCTCGACCGTCTCCTACTGCACCAGAACCGCGTGGCCCATGTGCACCCGCGATGCCCTT  
 CCGTGACCTTGGCCGCTCATGACACTCTATCTGTTTGCCAAACATCTATCAGCGCTGCCCA  
 CTGAGGCCCTGGCCCCCTCGCTGCGTGCCTGCGTACCTGAGGCTCAACGACAACCCCTGGGTG  
 TGTGACTGCCGGGCACGCCCCACTCTGGGCTGGCTGCAGAAGTTCGCGCGCTCCTCTCCGA  
 GGTGCCCTGCAGCCTCCCGCAACGCTGGCTGGCCGTGACCTCAAACGCTTAGCTGCCAATG  
 ACCTGCAGGGCTGCGCTGTGGCCACCGGCCCTTACCATCCCATCTGACCCGCGAGGCCACC  
 GATGAGGAGCCGTGGGGCTTCCCAAGTCTGCCAGCCAGATGCCGCTGACAAGGCCCTCAGT  
 ACTGGAGCCTGGAAGACCAGCTTCGGCAGGCAATGCGCTGAAGGGACGCGTGCCTGCCCGGTG  
 ACAGCCCGCGGGCAACGGCTCTGGCCACGGCACATCAATGACTCACCCCTTTGGGACTCTG  
 CCTGGCTCTGCTGAGCCCCGCTCACTGCAGTGCAGCGCCGAGGGCTCCGAGCCACAGGGTT  
 CCCACCTCGGGCCCTCGCGGAGGCCAGGCTGTTACGCAAGAACCGACCCGCGAGCCACT  
 GCCGTCTGGGCCAGGCAGGCAGCGGGGGTGGCGGGACTGGTGACTCAGAAGGCTCAGGTGCC  
 CTACCCAGCCTCACTGCAGCCTCACCCCCCTGGGCTGGCGCTGGTGCTGTGGACAGTGCT  
 TGGGCCCTGCTGACCCCCAGCGGACACAAGAGCGTGCTCAGCAGCCAGGTGTGTGTACATAC  
 GGGGTCTCTCTCCACGCGCCCAAGCCAGCCGGCGGGCCGACCCGTGGGGCAGGCCAGGCCAG  
 GTCCCTCCCTGATGAGCGCTGCCGCCGCCACCCCATCTCCACCCCATCATGTTTACAGGG  
 TTCGGCGCGCAGGTTTGTTCAGAAACGCGCCTCCACCCAGATCGCGGTATATAGAGATAT  
 GAATTTTATTTTACTTGTGTAATAATATCGGACGAGTGAATAAAGAGCTCTTTTCTTAA  
 AAAA

10017081-102401

## **FIGURE 161**

></usr/seqdb2/sst/DNA/Dnaseqs.min/ss.DNA44184  
><subunit 1 of 1, 473 aa, 1 stop  
><MW: 50708, pI: 9.28, NX(S/T): 6  
MKRASAGGSRL LAWLWLQAWQVAAPCPGACVCYNEPKVTTSCPQQGLQAVPVGIPAA SQRI  
FLHGNRI SHVPAASFRACRNLTILWLHSNVLARIDAAFTGLALLEQLDLSDNAQLRSVDP A  
TFHGLGR LHTLHLDRCGLQELGPGLFRGLAALQYLYLQDNALQALPDDTFRDLGNLTHLFLH  
GNRISSVPERAFRGLHSLDRLLLHQNRVAHVHPHAFRDLGRMLTYLYFANNLSALPTEALAP  
LRLQYLR LNDNFWVCD CRARPLWAWLQKFRGSSSEVP CSLPQRLAGRDLKRLAANDLQGCA  
VATGPYHPIWTGRATDEEPLGLPKCCQPDAA DKASVLEPGRPASAGNALKGRVPPGDSPPGN  
GSGPRHINDSPFGTLPGSAEPPLTAVRPEGSEPPGFPTSGPRRRPGCSRKNRTRSHCRLGQA  
GSGGGGTG DSEGS GALPSLTCSLTPLGLALVLWTVLGPC

### **Important features:**

#### **Signal peptide:**

amino acids 1-26

#### **Leucine zipper pattern.**

amino acids 135-156

#### **Glycosaminoglycan attachment site.**

amino acids 436-439

#### **N-glycosylation site.**

amino acids 82-85, 179-183, 237-240, 372-375 and 423-426

#### **WFC domain**

amino acids 411-425

# FIGURE 162

GGAAGTCCACGGGGAGCTTGGATGCCAAAGGAGGACGGCTGGGTCTCTGGAGAGGACTAC  
 TCACCTGGCATATTTCTGAGGTATCTGTAGAATAACCCAGCCTCAGATACTGGGGACTTTAC  
 AGTCCCAAGAAACCGTCTCCAGGAAGCTGAATCCAGCAAGAAACAATGGAGGCCAGCGGGA  
 AGCTCATTTGCAGACAAAGGCAAGTCTCTTTTTCTCTTTCTCCTTTTGGGCTTATCTCTGGCG  
 GGCCTGGCGGAACCTAGAAGCTATTTCTGTGGTGGAGGAAACTGAGGGCAGCTCCTTTGTAC  
 CAATTTAGCAAAAGGACCTGGGTCTGGAGCAGAGGGAATTTCCAGGCCGGGGGGTTAGGGTTG  
 TTTCCAGAGGGAACAAACTACATTTGCAGCTCAATCAGGAGACCGCGGATTTGTTGCTAAAT  
 GAGAAATTTGGACCGTGAGGATCTGTGCGGTACACAGAGCCCTGTGTCTACGTTTCCAAGT  
 GTTGTCTAGAGAGTCCCTTCGAGTTTTTTCAAAGCTGAGCTGCAAGTAAATAGACATAAACCGACC  
 ACTCTCCAGTATTTCTGGACAAACAAATGTTGGTGAAAGTATCAGAGAGCAGTCTCTCTGGG  
 ACTACGTTTCTCTGAAGAATGCCGAAGACTTAGATGTAGGCCAAACAAATATTGAGAACTA  
 TATAATCAGCCCCAATCTCTATTTTTCGGGTCTCACCCGCAACCGCAGTGATGGCAGGAAAT  
 ACCGAGAGCTGGTGCTGGACAAAGCGCTGGACCGAGAGGAAGAAGCTGAGCTCAGGTTAAACA  
 CTCACAGCACTGGATGGTGGCTCTCCGCCAGATCTGGCACTGCTCAGGCTACATCTCGAAGT  
 CCTGGATGTCAACGATAATGCCCTCGAATTTGAGCAGCCTTTCTATAGAGTGCAGATCTCTG  
 AGGACAGTCCGGTAGGCTCTCTGTTTGTGAAGTCTCTGCCACGGATGTAGACACAGGAGTC  
 AACGGAGAGATTTCTTATTTCACTTTTCCAAGCTTCAGAAAGAGATTGGCAAAACCTTTAAGAT  
 CAATCCCTTGCAGAGGAGAAATTTGAACATAAAACCAACTCGATTTGCAAAACCTTCAGTCTCT  
 ATGAAGTCAATATTGAGGCAAGAGATGCTGGAACTTTCTGGAAATGACCGGTTCTGTATT  
 CAAGTGATAGATGTGAACGACCATGCCCCAGAAAGTTACCATGTCTGCATTTACCAGCCCAAT  
 ACCTGAGAACCGCCCTGAAACTGTGGTTGCATTTTTCAGTGTTCAGATCTTGATTTCAGGAG  
 AAAATGGGAAAAATTAGTTGCTCCATTTCAGGAGGATCTACCTTCTCTGAAATCCGCGGAA  
 AACTTTTACACCTTACTAACGGAGAGACCACTAGACAGAGAAACGAGAGCGGAATACAACAT  
 CACTATCACTGTCTACTGACTTGGGGACCCCTATGCTGATAACACGACTCAATATGACCGTGC  
 TGATCGCCGATGTCAATGACAACGCTCCGCGCTTACCCTAACCTCTACACCTGTTCTGTC  
 CGCGAGAACAAACAGCCCCGCTGACATCCGACGCTCAGCGCTACAGACAGAGACTCAGG  
 CACCAACGCCCGAGGTCACTACTGCTGTGCCGCCAGGACCCGCACTGCCCCCTCACAT  
 CCCTGGTCTCCATCAACGCGGACAAACGCCACCTGTTTCGCCCTCAGGTCTCTGGACTACAG  
 GCCCTGACGGGGTTTCAGTTCCGCGTGGGCGCTTCAGACCACGGCTCCCCGGCGCTGAGCAG  
 CGAGGCGCTGGTGCGGTGTGGTGTGGACGCCAACGACAACTCGCCCTTCGTGCTGTACC  
 CGCTGCAGAACGGCTCCGCGCTTGCACCGAGCTGGTGCCCGGGCGGCGAGCCGGGCTAC  
 CTGGTGACCAAGGTGGTGGCGGTGGACGGCGACTCGGGCCAGACCGCTGGCTGTGCTACCA  
 GCTGCTCAAGGCCACGGAGCTCGGTCTGTTTCGGCGTGTGGGCGCAATAGGCGAGGTGCGCA  
 CCGCCAGGCTGCTGAGCGAGCGGACGCGGCCAAGCACAGGCTGGTGGTGTGGTCAAGGAC  
 AATGCGAGCTCCGCGCTCGGCCACCGCCAAGCTGCCTGCTCCTGGTGGACGGCTTCTC  
 CCAGCCCTACCTGCTCTCCCGGAGGCGGCCCGACCCAGGCCAGGCCAGCTGCTGCTACCG  
 TCTACCTGGTGGTGGCGGTGGCTCGGTGCTCTTCGCTCTTCTCTTTTTCGGTGCTCTGTTT  
 GTGGCGGTGCGCTGTGTAGGAGGAGCAGGGCGGCTCGGTGGGTGCTGCTGGTGGCGGA  
 GGGCCCCCTTCAGGGCATCTTTGTGGACATGAGCGGCACAGGACCTTATCCAGAGCTACC  
 AGTATGAGGTGTGCTGGCAGGAGCTCAGGACCAATGAGTTCAAGTCTCTGGAAGCCGAT  
 ATCCCAACTCTCCCTCCCAAGTGCCCTGGGAAAGAAATACAAGGAAATTTACCTTCCCCAA  
 TAACTTTGGTTCAATATTCAGTGAACCATAGTTGACTTTTACATTTCCATAGGTATTTTATTT  
 TGTGGCATTTCCATGCCAATGTTTATTTCCCCAAATTTGTGTGTATGTGAATATTGTACGGAT  
 TTACTCTTGATTTTCTCATGTTCTTTCTCCCTTTGTTTTAAAGTGAACATTTACCTTTATT  
 CCTGGTTCTT

10017001-102401

## **FIGURE 163**

</usr/seqdb2/sst/DNA/Dnaseqs.min/ss.DNA48314  
<subunit 1 of 1, 798 aa, 1 stop  
<MW: 87552, pI: 4.84, NX(S/T): 5  
MEASGKLICRQRQVLFSLFLLGLSLAGAAEPRSYVVEETEGSSSVFTNLAKDLGLEQREFSR  
RGVRVVS RGNKLHLQNLQETADLLNEKLDREDLCGHTEPCVLR FQVLLES PFEFFQ AELQV  
IDINDHSPVFLDKQMLVKVSESSPPGTTFFPLKNAEDLDVGQNNIENYIISPNSYFRVLTRKR  
SDGRKYPELVLDKALDREEEAELRLTLTALDGGSPPRSGTAQVYIEVLDVNDNAPEFEQPFY  
RVQISEDSPVGFLVVKVSATD VDTGVNGEISYSLFQASEEIGKTFKINPLTGEIELKKQLDF  
EKLQSYEVNIEARDAGTFSGKCTVLIQVIDVNDHAPEVTMSAFTSPIENAPETVVALFSVS  
DLDSGGENKISCISQEDLPFLKSAENFYTLTERPLDRESRAEYNITITVTDLGTPLITQ  
LNMTVLIADVNDNAPAFQTSTYTLFVRENNSPALHIRSVSATDRDSGTNAQVTYSLPPQDP  
HLPLTSLVSINADNGHLFALRSLDYEALQGQFRVGASDHGSPALSSEALVRVVVLDANDNS  
PFVLYPQNGSAPCTELVPRAAEPGYLVTKVVAVDGDSGQNAWLSYQLLKATELGFGVWAH  
NGEVRTARLLSERDAAKHRLVVLVKDNGEPPRSATATLHVLLVDGFSQPYLPLPEAAPTQAQ  
ADLLTVYLVVALASVSSLFLFVSVLLFVAVRLCRRSRAASVGRCLVPEGPLPGHLDMSGTRT  
LSQSYQYEVCLAGGSGTNEFKFLKPIIPNFPQCPGKEIQGNSTFPNNFGFNIQ

**Important features:**

**Signal peptide:**

amino acids 1-26

**Transmembrane domain:**

amino acids 685-712

**Cadherins extracellular repeated domain signature.**

amino acids 122-132, 231-241, 336-346, 439-449 and 549-559

**ATP/GTP-binding site motif A (P-loop).**

amino acids 285-292

**N-glycosylation site.**

amino acids 418-421, 436-439, 567-570 and 786-789

## FIGURE 164

ACCCACGCGTCCGCCCACGCGTCCGCCCACGCGTCCGCCCACGCGTAGCCGTGC  
GCCGATTGCTCTCGGCTGGGCAATGCTCCGGCTGCCGTCGACGACCGCCCCGCGTCAT  
CGGGCTCCTCGGCTGGTGGCAAGTATTGCTGTGGGTGCTGGGACTTCCCGTCCGCGGCGTGG  
AGGTTGCAGAGGAAAGTGGTCGCTTATGGTCAGAGGAGCAGCCTGCTCACCCTCTCCAGGTG  
GGGGCTGTGTACCTGGGTGAGGAGGAGCTCTGCATGACCCGATGGGCCAGGACAGGGCAGC  
AGAAGAGGCCAATGCGGTGCTGGGGCTGGACACCCAAGGCGATCACATGGTGATGCTGTCGTG  
TGATTCTGGGGAAGCTGAGGACAAAGTGAGTTCAGAGCCTAGCGGCGTCACCTGTGGTGTCT  
GGAGGAGCGGAGGACTCAAGGTGCAACGTCCGAGAGAGCCTTTTCTCTCTGGATGGCGCTGG  
AGCACACTTCCCTGACAGAGAAGAGGAGTATTACACAGAGCCAGAAGTGGCGGAATCTGACG  
CAGCCCCGACAGAGGACTCCAATAACACTGAAAGTCTGAAATCCCCAAAGGTGAACTGTGAG  
GAGAGAAACATTACAGGATTAGAAAATTTCACTCTGAAAATTTTAAATATGTCACAGGACCT  
TATGGATTTTCTGAACCCAAACGGTAGTGACTGTACTCTAGTCTGTTTACACCCCGTGGT  
GCCGCTTTTCTGCCAGTTTGGCCCTCACTTTAACTCTCTGCCCCGGGCATTTCCAGCTCTT  
CACTTTTGGCACTGGATGCATCTCAGCACAGCAGCCTTTCTACCAGGTTTGGCACCCTAGC  
TGTTCCCTAATATTTTATTATTTCAGGAGCTAAACCAATGGCCAGATTTAATCATACAGATC  
GAACACTGGAAACACTGAAAATCTTCATTTTAAATCAGACAGGTATAGAAGCCAAGAAGAAT  
GTGGTGGTAACTCAAGCCGACCAATAGGCCCTCTTCCAGCACTTTGATAAAAAGTGTGGA  
CTGGTGTCTGTATTTTCTTATCTTTTAAATAGTTTTATTATGTATGCTACCATTTCGAA  
CTGAGAGTATTCGGTGGCTAATTCCAGGACAAGAGCAGGAACATGTGGAGTAGTGATGGTCT  
GAAAGAAGTTGGAAGAGGAACTTCAATCCTTCGTTTCAGAAATTAGTGCTACAGTTTCATA  
CATTTTCTCCAGTGACGTGTTGACTTGAACTTCAGGCAGATTTAAAGAATCATTTGTTGAA  
CAACTGAATGTATAAAAAAATTATAAACTGGTGTTTTAACTAGTATTGCAATAAGCAAATGC  
AAAAATATTCAATAG

10077081-102401

## FIGURE 165

></usr/seqdb2/sst/DNA/Dnaseqs.min/ss.DNA48333  
><subunit 1 of 1, 360 aa, 1 stop  
><MW: 39885, pI: 4.79, NX(S/T): 7  
MVPAAGRPRPRVMRLLGWWQVLLWVLGLPVRGVEVAEESGRLWSEEQPAHPLQVGAVYLGEE  
ELLHDFPMGQDRAAEAEANAVLGLDTQGDHVMVLSVIPGEAEDKVSSEPSGVTCGAGGAEDSRC  
NVRESLFLSLDGAGAHFPDREEEYYTEPEVAESDAAPTEDSNNTESLKSFPKVNCEERNITGLE  
NFTLKILNMSQDLMDFLNPNGSDCTLVLFYTPWCRFSASLAPHFNSLPRAFPAHFALDAS  
QHSSLSTRFGTVAVPNILLFQGAKPMARFNHTDRTLTLTKIFIFNQTGIEAKKNVVVTQADQ  
IGPLPSTLIKSVDWLLVFSLFFLISFIMYATIRTESIRWLIPGQEQEHVE

### **Important features:**

#### **Signal peptide:**

amino acids 1-25

#### **Transmembrane domain:**

amino acids 321-340

#### **Homologous region to dilsufide isomerase**

amino acids 212-302

#### **N-glycosylation site.**

amino acids 165-168, 181-184, 187-190, 194-197, 206-209, 278-281  
and 293-296

#### **Thioredoxin domain**

amino acids 211-227

CCCGGCTCCGCTCCCTCTGCCCCCTCGGGGTCGCGCGCCACGATGCTGCAGGGCCCTGGCT  
CGCTGCTGCTGCTCTTCTCGCTCGCACTGCTGCTGGGCTCGCGCGCGGGCTCTTCTC  
TTTGGCCAGCCGACTTCTCTACAAGCGCAGCAATTGCAAGCCATCCGGTCAACCTGCA  
GCTGTGCCACGGCATCGAATACCAGAACTGCGGCTGCCAACCTGCTGCGGCACGAGACC  
TGAAGGAGGTGCTGGAGCAGGCCGCGCTTGGATCCCGCTGGTATGAAGCAGTGCCACCG  
GACACCAAGAAGTCTCTGTGCTCGCTCTTCGCCCCGCTGTGCTCGATGACCTAGACGAGAC  
CATCCAGCCATGCCACTCGCTCTGCGTGCAGGTGAAGGACCGCTGCGCCCGGTATGTCCG  
CCTTCGGCTTCCCTTGCCCGACATGCTTGAGTGCAGCCGTTTCCCCAGGACAACGACCTT  
TGCACTCCCGCTGCTAGCAGCGACCACTCTGCCAGCAGCAGGAGAGCTCAAAAGTATG  
TGAAGCCTGCAAAAATAAAAAATGATGATGACAACGACATAATGGAACGCTTTGTAAAAATG  
ATTTTGCACGTGAAAATAAAAGTGAAGGAGATAACCTACATCAACCGAGATACCAAATCATC  
CTGGAGACCAAGAGCAAGACCATTTACAAGCTGAACGGTGTGTCCGAAAGGGACCTGAAGAA  
ATCGGTGCTGTGGCTCAAAGACAGCTTGCACTGTGAGGAGATGAACGACATCAACG  
CGCCTATCTGGTTCATGGGACAGAAACAGGGTGGGGAGCTGGTGATCACTCGGTGAAGCGG  
TGGCAGAAGGGGCGAGAGAGTTCAAGCGCATCTCCCGCAGCATCCGACGCTGCAGTGCTA  
GTCCCGGCATCCTGATGGCTCCGACAGGCCTGTCCAGAGCACGGCTGACCAATTTCTGTCTC  
GGGATCTCAGCTCCCGTTCCCAAGCACACTCTAGCTGCTCCAGTCTCAGCCTGGGAGCT  
TCCCCCTGCTTTTTCACGCTTTTGACATCCCCAGCATTTCTCTGAGTTATAAGGCCACAGGAGTG  
GATAGCTGTTTTCACCTAAAGGAAAAGCCACCCGAATCTTGTAGAAATATTCAAACATAA  
AAATCATGAATATTTTAA

## **FIGURE 167**

```
></usr/seqdb2/sst/DNA/Dnaseqs.min/ss.DNA50920
><subunit 1 of 1, 295 aa, 1 stop
><MW: 33518, pI: 7.74, NX(S/T): 0
MLQGPGSLLLLFLASHCCLGSARGLFLFGQPDFSYKRSNCKPIPVNLQLCHGIEYQNMRLPN
LLGHETMKEVLEQAGAWIPLVMKQCHPDTKKFLCSLFAPVCLDDLDETIQPCHSLCVQVKDR
CAPVMSAFGFPWPDMLECDRFPQDNDLCIPLASSDHLLPATEEAPKVCEACKNKNDNDNDIM
ETLCKNDFALKIKVKEITYINRDTKIILETKSKTIYKLVGVSERDLKKSVLWLKDSLQCTCE
EMNDINAPYLVMGQKQGGELVITSVKRWQKGQREFKRISRIRKLQC
```

**Important features:**

**Signal peptide:**

amino acids 1-20

**Cysteine rich domain, homologous to frizzled N terminus**

amino acids 6-153

## FIGURE 168

GTGGAGGCCGCCGACGATGCGGGGCCGACGGAGGCCGAGACGGGGTTGGCCGAGCCCCGGG  
CCCTGTGCGCGCAGCGGGGCCACCGACCTACGCGCGCCGCTGGGTGTTCTGTCTCGCGATC  
AGCCTGCTCAACTGCTCCAACGCCACGCTGTGGCTCAGCTTTGCACCTGTGGCTGACGTCA  
TGCTGAGGACTTGGTCCTGTCCATGGAGCAGATCAACTGGCTGTCACTGGTCTACCTCGTGG  
TATCCACCCCATTTGGCGTGGCGGCCATCTGGATCCTGGACTCCGTCGGGCTCCGTGCGGCG  
ACCATCCTGGGTGCGTGGCTGAACCTTTGCCGGGAGTGTGTACGCATGGTGCCTGCATGGT  
TGTGGGACCCAAAACCATTTGCCCTTCTCATGGGTGGCCAGAGCCTCTGTGCCCTTGCCC  
AGAGCCTGGTCATCTTCTCTCCAGCCAAGCTGGCTGCCTTGTGGTTCACAGAGCACCAGCGA  
GCCACGGCCAACATGCTCGCCACCATGTGCAACCCCTTGGGCGTCCTTGTGGCCAATGTGCT  
GTCCCTGTGCTGGTCAAGAAGGGTGAGGACATTCGGTTAATGTCCGTGTCTATACCATCC  
CTGCTGGCGTCGTCTGCCTGCTGTCCACCATCTGCCTGTGGGAGAGTGTGCCCCCACCCTCG  
CCCTCTGCGGGGCTGCCAGCTCCACCTCAGAGAAGTTCTCGGATGGGCTCAAGCTGCAGCT  
CATGTGGAACAAGCCCTATGTCATCCTGGCTGTGTGCTTGGGGGAATGATCGGGATCTCTG  
CCAGCTTCTCAGCCCTCTCGGAGCAGATCCTCTGTGCAAGCGGCCACTCCAGTGGGTTTTCC  
GGCCTCTGTGGCGCTCTCTTCATCACGTTTGGGATCCTGGGGGCACTGGCTCTCGGCCCTA  
TGTGGACCGGACCAAGCACTTCACTGAGGCCACCAAGATTGGCCTGTGCCTGTTCTCTCTGG  
CCTGCGTGCCTTTTGCCCTGGTGTCCAGCTGCAGGGACAGACCTTGCCCTGGCTGCCACC  
TGCTCGCTGCTCGGGCTGTTTGGCTTCTCGGTGGGCCCCGTGGCCATGGAGTTGGCGGTGCGA  
GTGTTCTCTTCCCGTGGGGGAGGGGGCTGCCACAGGCATGATCTTTGTGCTGGGGCAGGCCG  
AGGGAATACTCATCATGCTGGCAATGACGGCACTGACTGTGCGACGCTCGGAGCCGTCCTTG  
TCCACCTGCCAGCAGGGGGAGGATCCACTTGACTGGAAGTGTCTCTGCTGCTGATGGCCGG  
CCTGTGCACCTTCTTACGCTGCATCCTGGCGGTCTTCTTCCACACCCCATACGGGCGCCTGC  
AGGCCGAGTCTGGGGAGCCCCCTCCACCCGTAACGCCGTGGGCGGCGCAGACTCAGGGCCG  
GGTGTGGACCGAGGGGGAGCAGGAAGGGCTGGGGTCTGGGGCCAGCACGGCGACTCCGGA  
GTGCACGGCGAGGGGGGCTCGCTAGAGGACCCAGAGGGCCCGGAGCCCCACCCAGCCT  
GCCACCGAGCGACTCCCGTGCACAAGGCCACGACGACCGACGCGCCCTCCGCCCCGCG  
AGACTCGCAGGCAGGGTCCAAGCGTCCAGGTTTATTGACCCGGCTGGGTCTCACTCCTCCTT  
CTCCTCCCGTGGGTGATCACGTAGCTGAGCGCCTTGTAGTCCAGGTTGCGCCGACATCGA  
TGGAGGCGAACTGGAACATCTGGTCCACCTGCGGGCGGGGGCGAAAGGGCTCCTTGGGGCT  
CCGGGAGCGAATTACAAGCGCGCACCTGAAAA

1001-1001-1001-1001

## FIGURE 169

></usr/seqdb2/sst/DNA/Dnaseqs.min/ss.DNA50988

><subunit 1 of 1, 560 aa, 1 stop

><MW: 58427, pI: 6.86, NX(S/T): 2

MAGPTEAETGLAEPRALCAQGRHRTYARRWVFLLAISLLNCSNATLWLSFAPVADVIAEDLV  
LSMEQINWLSLVLYLVSTPFGVAAIWILDSVGLRAATILGAWLNFAGSVLRMVPCMVVGTQN  
PFAFLMGGQSLCALAQSLVIFSPAKLAALWFPEHQATANMLATMSNPLGVLVANVLSVPLV  
KKGEDIPLMLGVYTI PAGVVCLLSTICLWESVPPTPPSAGAASSTSEKFLDGLKLQLMWNKA  
YVILAVCLGGMIGISASFSALEQILCASGHSSGFSGLCGALFITFGILGALALGPYVDRTK  
HFTEATKIGLCFLSLACVPFALVSQLQGQTLALAATCSLLGLFGFSVGPVAMELAVECSFPV  
GEGAATGMIFVLGQQAEGILIMLMTALT VRRSEPSLSTCQQGEDPLDWTVSLLLMAGLCTFF  
SCILAVFFHTPYRRLQAESGEPSTRNAVGGADSGPGVDRGGAGRAGVLGPSTATPECTARG  
ASLEDPRGPGSPHPACHRATPRAQGPAA TDAPSRPGRLAGRVQASRFIDPAGSHSSFSSPWVIT

### **Important features:**

#### **Signal peptide:**

amino acids 1-44

#### **Transmembrane domains:**

amino acids 61-79, 98-112, 126-146, 169-182, 201-215, 248-268,  
280-300, 318-337, 341-357, 375-387, 420-441

#### **N-glycosylation site.**

amino acids 40-43 and 43-46

#### **Glycosaminoglycan attachment site.**

amino acids 468-471

# **FIGURE 170**

GTCCACATCTCTGCTCAACTGGGTGAGGTCCCTCTTAGACAGCTCTTTGTCCATCATTTTGTCTGAAGTGGACCAAC  
TAGTTCGCCAGTAGGGGCTCTCCCTGGCAATTCTTGATCGCGCTTTGGACATCTCAGATCGCTTCCAATGAAGA  
TGGCCTTGCCCTTGGGGTCTGCTGTGTTTCAATAATCACTAACTATGGGCAAGGCTTTGGCCGGCAGCTCTGGGGG  
AAGGAGCACTGGGGCTGATCAAGCCATCCAGGAAACACTGGAGGACTTCTCCAGCCTTGAAAGACTCTAGTGTGTT  
TCTGAATCTAGCCCACTTTGGCGGTAAAGCATGATGCAACTTCTGCAACTTCTGCTGGGGCTTTTGGGGCCAGGTGG  
CTCTTATTCTTTTAGGGGATTTGTCAAGGAGTGACCACTCTCACGGTGAATAACAAGTGTCTAGAGGAAGTCC  
ATCTGTGACATGTATCGGGAAGCTGTCCAGGAATCTGGGCGGAGGAGGAGGCGAGGCAAGTCTGGGGCGCCTGT  
CCAGGTTGTTGTCAGCTGCTCTCAGGCGCTCCCAATTCAAGTGGATCTGAGGAAGGCTGCTCAGCAGAGGCGAGGCG  
GCTGGATCGAGAGAGCTGTGCCAGCAGTGGGATCCCTGCTGGTTCTTTGATGTGCTTGGCACAAGGGGATTT  
GGCTCTGATTCATGTGGAGATCCAAGTGTGACATCAATGACCAACAGGCAAGGTTTCCCAAGGCGAGGAGGAGA  
GCTGGAAATCTCTGAGAGCGCTCTCTGCGAACCAGGATCCCTCTGGACAGGACTCTTGACCCAGACACAGCGCCC  
TAACACCTCTGACACCTCACTCTGTCTCCAGTGAGCACTTTGCTTGTGATGTCAATGTGGGCGCTGATGAGAC  
CAACATGCGAAGCTCATAGTGGTGAAGGAGCTGGACAGGGAATCCATTCAATTTTGTGATCTGGTGTAACTGC  
CTATGACAAATGGGAACCCCCCAAGTCAAGTACAGCTTGTGCAAGGTCAACGCTTGGACTCCAATGACAATAG  
CCCTGCGTTTGTCTGAGAGTTCACTGGCACTGGAAATCCAAGAAGATGCTGCACTGGTACGCTTCTCATAAAACCT  
GACGCGCACAGACCTTGAACAGGCCCCCAATGGGAGGTTGGAATCTTCTCATGTAAGCACATGCTCCAGAGGT  
CCGTGCAACTTTCAGTATTGATGTCACAGCAGGCGAGGTCACTTGTGCTGCACTCTAGACTTGAATGAAAAGACCC  
TGCTCTCAGAGTGTGATGTTCAAGCAAGGACCTGGGTCCCAATCTCATCCAGGCCATTCGAAGTTTCTCATCAA  
GGTTCTGAGATGTTCAATGATGTCACCAAGCTCCAGCTCAGATGGGCTCCAGCAGCTCACTGGTGTCAAGAGC  
TCTTCCAGAGGACAGCTTTTATTGCTCTTGTCTAGGAGATGACTTGGATTCAAGGACCAATGGTTTGGTCCAGTGC  
CTGGCTGCGCCAGAGCTGGGCGACTCTCAGGCTGAAAAGAACTAATGGCAACATACATGTTGTCTAACCAATGC  
CACATCTGACAGAGAGCAGTGGGCCAAATATACCTCACTCTGTAGGCCAACAGGCAAGGACTCCAGCGCTTATC  
AGCCAGAAACAGCTCAGCAATTCAGATCAGTGACATCAACGACATGCACTTGTGTTGAGAAAGCAGGTATGA  
AGTCTCCAGCGGGAAACAACTTACCTCTCTTCACTCTATTACCATCAAGGCTCATGATGAGATCTGGGCACT  
TAATGAAAAGTCTCATACCGATCTCAGGACTCCCACTTGTCTCTAGTACTATGCTCAACACAGAGGA  
GGTCACTGCTCAGAGGTCTCACTGAATCTGAAGAGATGGCGGCTTTGATGTTCCAGGTGATCGACAGGAGCAGCGG  
GCAACCACTTGTGATCGACTGTCTCTGTGGGTGAGCTTCTGATGCAATGATGATGTTGCCAGGAGGTGGT  
CCAGCTGTGCTCAGGATGGAAGAGCAGGCTCTCCGTCTTGTGAATGCTCCACAGGCAACCTGCTGCGGTGCC  
CATCGAGACTCCCAATGGCTTGGGCGGAGGAGTGCAGACTCGCGGGCAATGAGAGGCCCTCTACAGACTCCGCAATGG  
AAATGAGGCCCACTCTTATCTTCAACCTCATACGGGCGAGTGTTCGTCAATGTCAACATGACAGCAGCT  
CATTTGGGATGTAGTGGAGCTGGAGATAGTAGTAGAGGACCCAGGGAAGCCCCCTTACAGACCCCGAGCCTGT  
GAGGCTCATGTTTGTCAACAGTGTGGACCACTGAGGAGCTCAGCGCGCAAGCTTGGGGCTTGAGCATGTGAT  
GCTGACGGGTGATCTGCTGGCTGTACTGTTGGGCTCTTGGGTTGATCTGCTGATGCTTCTCATGTCTCACTGCGG  
GACAGAAAAGAGGACACAGGGCTTCAACTGTGGGAGGCGAGTCCACTTACCGGACAGGCCCCAAGAGGCC  
CCAGAAACACATTCAGAAGGCAGACATCCACTCTGTGCTGTGCTCAGGGGTGAGGCTGAGCTTGTGAAGT  
CGGGCAGTCCCAAGAAGTGTGGACAGGAGGCGATGATGGAAGCAGGCTGGGAGCTCCCTGCTGCGAGGCCCTT  
CCACTTCAACCGCCCTGTGACAGGCGAGATCTCCCTTTCACCATCCAGGCGAGGAATGCTTCCGGGAGGAACCTGAACCTTCC  
CGAGCCCGAGCTGCGCACCGCCAGCCAGTTCAGGCTTCAAGGTTGTCAGGCGAGCCCAAGGAGGGTGGC  
TGAGACAGAGGCGAGTGGGAAGGCCACAGAGGCGCCAGGCTCTCTGCAACCTGAGACCGGCGAGCATCT  
CAATGGCAAGTGTCCCTTGAGAAAGAACTCAGGCGCCGTCAGTCTGCGAGGCTGAGTCTGAGTCTGTGTCG  
TGCTCTGCGCGAGCGGAACCCGCTGGAGGAGCTCACTTGGATCTTCTCTGCTGTGACAAATCTCCAGCTGCT  
GTCTTCTGCTCATCAGGCGCACTCTCAGGCCAAACCAAAACACAGGAGAAATAGTACTTGGCCACAGGAGG  
CAGCAGGATGTCAATCCAGACACAGATGGCCCAAGTGAAGGCTGAGGCGCAGACAGCCAGAACAGGAGGA  
AGGCGTTTGGATCTTGAAGAGGACCTCTCTGTGAAGCACTGCTGAAGAGAGAGCTTCAAGTCTGTGTCGACCC  
CAGCAGAGTCTGGCCCTGGAACCGGCTGAGCGGCCCTGACCGGCTGGATGGCGAGACTCTTCTTGCCCTTAC  
CACTCACTACGTCGACAACTGTGATCTCCCGGATGCTGACGCAAGGAGGAGCGGAGGACTTCCAGAGCTTGG  
CAAGGCGAGAGGACAGAGCTGAGCCCAACAGGCAAGGAGTGGCCAGCACTTGTCTCGGAGATGAGCTCACT  
GCTGGAGATGCTGTGAGACAGGCTCCAGCATGCCGTGGAGGCGGCTTCCAGGCGGCTTGGCGGCTCTCTGCT  
CTGCGGAGGACCTCAGTTTGAATCTTGGCCACAGTGCAGCTCAGGCAAGTGAAGTGAAGGAGCCAGGAGT  
AAGAGCGGGGACTGAGGCGAAGGACAGGAGCAGCAGCAGCAGGAGTCTGAGACATACCTCAGACGCT  
CTGGATCTCAAGAACCGGGGCTGAGGATCTGTGGACAGAGCTGTTTCAAAATCTTGAATCACTAGCTAG  
CGGCGGCTGAGAACTTTAGGTGACTGATGCTACCCCAAGAGGAGGACAGGCGCCAGGACTAACAGCTGAC  
TGCCCAAGGAGGCGCTTGTGAAGCAGCTGAGTCTTTGGAGGACAGGACGTTTGTGGTGTGAGATAAGTGT  
TCTCTGCAAAACATATGTGGAGCAAAAGGCTAGTCTCTGCGAGAACGATGTCAGGAGTACAGCAGG  
AAAGGTGGCTCTTCTGGGTAGCAGGAGTCAAGGCGCTGACCTGGGCGGTGCCAGGAAATGCTCTTGAACCTAT

1007081.1001

## FIGURE 171

</usr/seqdb2/sst/DNA/Dnaseqs.min/ss.DNA48331  
<subunit 1 of 1, 1184 aa, 1 stop  
<MW: 129022, pI: 5.20, NX(S/T): 5  
MMQLQLQLLGLLGGPGGYLFLLGDCQEVTTLTVKYQVSEEVPSGTVIGKLSQELGREERRRQA  
GAAFQVLQLPQALPIQVDSEGLLSTGRRLDREQLCRQWDPCLVSFVDVLATGDLALIHVEIQ  
VLDINDHQPRFPKGEQELEISASLRTRIPLDRALDPPDTGPNLTHTYTLSPSEHFALDVI  
GPDETKHAELIVVKELDREIHSFFDLVLTAIDNGNPPKSGTSLVKVNVLDSDNDSPAPFAESS  
LALEIQEDAAPGTLIKLTATDPDQGPNGEVEFFLSKHMPPPEVLDTFSIDAKTGQVILRRPL  
DYEKNPAYEVDVQARDLGNPIPAHCKVLIKVLDVNDNIPSIHVTWASQPSLVSEALPKDSF  
IALVMADDLDSGHNGLVHCWLSQELGHFRLKRTNGNTYMLLTNATLDREQWPKYTLTLAQD  
QGLQLSIAKKQLSIQISDINDNAPVFEKSRYEVSSTRENNLPSLHLITIKAHADLGLINGKVS  
YRIQDSPVAHLVAIDNSNTGEVTAQRSLNYEEMAGFEFQVIAEDSGQPMPLASSVSVVSLDA  
NDNAPEVVQPVLSDGKASLVLVNASTGHLLVPIETPNGLGPGAGTDTPPLATHSSRPFLTT  
IVARADSGANGEPLYSIRNGNEAHLFIINPHTGQLFVNVTNASSLIGSEWELEIVVEDQGS  
PPLQTRALLRVMTVSVDHLRDSARKPGALSMSMLTVICLAVLLGIFGLILALFMSICRTEK  
KDNRAYNCREAESTYRQPKRPQKHQKADIHLVPVLRGQAGEPCEVQSHKDVKEMAMEA  
GWDPCQLQAPPHLTPTLYRTLNRQNGQAPAESREVLQDVTNLLFNHPRQRNASRENLNLEP  
QPATGQPRSRPLKVAGSPTGRLAGDQGSEAPQRPASSATLRRQRHLNGKVSPEKESGPRQ  
ILRSLVRLSVAFAERNPVEELTVDSPPVQQISQLLSLLHQGQFPKPNHRGNKYLAKEGGS  
RSAIPDTDGSPARAGGQTDPEQEEGPLDPEEDLSVKQLLEEELSSLLDPSTGLALDRLSAPD  
PAWMARLSLPLTNYRDNVISPDAAAATEEPRTFTQTFGKAEPALSPTGTRLASTFVSEMSSL  
LEMLELQRSSMPVEAASEALRRLSVCGRTLSLDLATSAAAGMKVQGDGPGKGTGTEGKSRGSS  
SSSRCL

### **Important features:**

#### **Signal peptide:**

amino acids 1-13

#### **Transmembrane domain:**

amino acids 719-739

#### **N-glycosylation site.**

amino acids 415-418, 582-585, 659-662, 662-665 and 857-860

#### **Cadherins extracellular repeated domain signature.**

amino acids 123-133, 232-242, 340-350, 448-458 and 553-563

## FIGURE 172

CGGACGCGTGGGCGGACGCGTGGGGGAGAGCCGCGAGTCCC GGCTGCAGCACCTGGGAGAAGG  
CAGACCGTGTGAGGGGGCTGTGGCCCCAGCGTGCTGTGGCTCGGGGAGTGGGAAGTGGAG  
GCAGGAGCCTTCCTTACACTTCGCCATGAGTTTCCTCATCGACTCCAGCATCATGATTACCT  
CCCAGATACTATTTTTTGGATTTGGGTGGCTTTTCTTCATGCGCCAATTGTTTTAAAGACTAT  
GAGATACGTAGTATGTTGTACAGGTGATCTTCCGTGACGTTTGCATTTTCTTGACCAT  
GTTTGAGCTCATCATCTTTGAAATCTTAGGAGTATTGAATAGCAGCTCCCGTTATTTTCACT  
GGAAATGAACCTGTGTGAATTTCTGCTGATCCTGGTTTTTCATGTGCGCTTTTACATTGGC  
TATTTTATTGTGAGCAATATCCGACTACTGCATAAAACAACGACTGCTTTTTCTGTCTCTT  
ATGGCTGACCTTTATGTATTTCTTCTGGAACCTAGGAGATCCCTTTCCCATTTCTCAGCCCAA  
AACATGGGATCTTATCCATAGAACAGCTCATCAGCCGGGTGGTGTGATTGGAGTGACTCTC  
ATGGCTCTTCTTTCTGGATTGGTGCTGTCAACTGCCCATACACTTACATGTCTTACTTCTCT  
CAGGAATGTGACTGACACGGATATTCTAGCCCTGGAACGGCGACTGCTGCAAAACCTGGATA  
TGATCATAAGCAAAAAGAAAGGATGGCAATGCGACGGAGAACATGTTCCAGAAGGGGGAA  
GTGCATAACAACCATCAGGTTTCTGGGGAATGATAAAAAGTGTTACCACCTCAGCATCAGG  
AAGTGAAAATCTTACTCTTATTCAACAGGAAGTGATGCTTTGGAAGAATTAAGCAGGCAGC  
TTTTTCTGGAACAGCTGATCTATATGCTACCAAGGAGAGAATAGAATACTCCAAAACCTTC  
AAGGGGAAATATTTTAATTTTCTGGTTACTTTTTCTCTATTACTGTGTTTGAAAAATTTT  
CATGGCTACCATCAATATTGTTTTGATCGAGTTGGGAAAAACGGATCCTGTCAACAGAGGCA  
TTGAGATCACTGTGAATTATCTGGGAATCCAATTGATGTGAAGTTTGGTCCCAACACATT  
TCCTTCATTCTGTGTGAATAATCATCGTCACATCCATCAGAGGATTGCTGATCACTCTTAC  
CAAGTTCTTTTATGCCATCTCTAGCAGTAAGTCTCCAATGTCAATTGTCTGTCTATTAGCAC  
AGATAATGGGCATGTACTTTGTCTCCTCTGTGCTGCTGATCCGAATGAGTATGCCTTTAGAA  
TACCGCACCATAAATCACTGAAGTCTTGGAGAACTGCAGTTCAACTTCTATCACCGTTGGTT  
TGATGTGATCTTCTGGTCAGCGCTCTCTAGCATACTCTTCTCTATTTTGGCTCACAAC  
AGGCACCAGAGAAGCAATGGCACCTTGAACTTAAGCCTACTACAGACTGTTAGAGGCCAGT  
GGTTTTCAAAATTTAGATATAAGAGGGGGGAAAAATGGAACAGGGCCTGACATTTTATAAAC  
AAACAAAATGCTATGGTAGCATTTTTACACCTTCATAGCATACTCCTTCCCCGTGAGTGATA  
CTATGACCATGAGTAGCATCAGCCAGAACATGAGAGGGAGAACTAACTCAAGACAATACTCA  
GCAGAGAGCATCCCGTGTGGATATGAGGCTGGTGTAGAGGCGGAGAGGCCAAGAACTAA  
AGGTGAAAAATACACTGGAACCTCTGGGGCAAGACATGTCTATGTTAGCTGAGCCAAACACGT  
AGGATTTCCGTTTTAAGGTTACATGGAAGGTTATAGCTTTGCCTTGAGATTGACTCATT  
AAAATCAGAGACTGTAACAAAAAAGGCGCGCGACTCTAGAGTCG  
ACCTGCAGAAGCTTGGCCGCATGGCCCAACTTGTTTATTGCAGCTTATAATG

## **FIGURE 173**

MSFLIDSSIMITSQILFFGFGWLFFMRQLFKDYEIRQYVVQVIFSVTFAFSCTMFELIIFBI  
LGVLNSSSRYPFHWMNLCVILLILVFMVPPYIGYFIVSNIRLLHKQRLLFSCLLWLTFFMYFF  
WKLGDPPFILSPKHGILSIEQLISRVGIVGVTLMALLSGFGAVNCPYTYMSYFLRNVTDTDI  
LALERLLQTMDMIISKKKRMAMARRTMFQKGEVHNKPSGFWGMIKSVTTSASGSENLTLIQ  
QEVDALEELSRQLFLETADLYATKERIEYSKTFKGKYFNFLGYFFSIYCVWKIFMATINIVF  
DRVGKTDPPVTRGIEITVNYLGIQFDVKFWSQHISFILVGIIIVTSIRGLLITLTKFFYAISS  
SKSSNVIVLLLAQIMGYFVSSVLLIRMSMPLEYRTIITEVLGELQFNFYHRWFDVIFLVSA  
LSSILFLYLAHKQAPEKQMAP

### **Important features:**

#### **Signal peptide:**

amino acids 1-23

#### **Potential transmembrane domains:**

amino acids 37-55, 81-102, 150-168, 288-311, 338-356, 375-398,  
425-444

#### **N-glycosylation sites.**

amino acids 67-70, 180-183 and 243-246

#### **Eukaryotic cobalamin-binding proteins**

amino acids 151-160

## FIGURE 174

CATGGGAAGTGGAGCCGGAGCCTTCCTTACACTCGCCATGAGTTTCCTCATCGACTCCAGCA  
TCATGATTACCTCCCNAGANACTATTTTTTGGATTGGGTGGCTTTTCTTCNGCGCCAATGTT  
TAAAGACTATGAGATACGTCAGTATGTTGTACNGGTGATCTTCTCCGTGACGTTTGCCATTT  
CTTGCACCATGTTTGAGCTCATCATCTTTGAAATCTTNGGAGTATTGAATAGCAGCTCCCGT  
TATTTTCACTGGAAAATGAACCTGTGTGTAATTCTGCTGATCCTGGTTNTCATGGTGCCTTT  
TTACATTGGCTATTTTATTGTGAGCAATATCCGACTACTGCATAACAACGACTGCTTTTT  
CCTGTCTCTTATGGCTGACCTTTATGTATTTCCAG



# FIGURE 176

CTCGCGCAGGGATCGTCCCATGCGCGGGGCTCGGAGCCGCGACCCCTTGGGGGGCCCTCCGGGATTGCTACCTTTT  
 TGGCTCCCTGCTCGTCCAGACTGCTCTTCTCAAGGGCTGTCGCTTCAATCTGAGAGTGTATGGGTGGCTTGGCGAA  
 GGAGGGCGAGCCAGGACGCTCTTGGCTTCTGTTGGCCCTGCACCGGAGTGTACGCCCGCCAGCCAGAGCTG  
 GCTGTGTGGTGGTGTCTCCAGGCGCTGCTCTTCTGGGCGAGGCGAATGCATCTGAGAGGCTCTTGTGCTTTG  
 CCGGTGTAGGCTGGAGGAGCTGACTGTCTACAGAGTGGACATCGACAGGAGGCTGATATGCAAAAGGAAAGCAA  
 GGAAGAACCACTGGTGGAGTCACTGTTCGAGCCAGGGGCTGGGGGCAAGATTGTACTCTGTGCAACCCGATA  
 TGAGGCAAGGCGAGGAGTGGACCAAGATCTGGAGAGCGCGGATATGATTTGGTGGCTTTGTGCTCAGCCAGGA  
 CTTGGCCATCCGGGATGAGTTGGATGGTGGGGAAATGGAAGTTCTGTGAGGGGAGCCGCCAAGGCCATGAACAATT  
 TGGGTTCTGCGCAGCAGGCGACAGCTGCGCTTCTCCCTTGATAGCCACTACTCTCTTTGGGGGCCAGGAAAC  
 CTATAATTGGAAGGCGACGCGCAGGGTGGAGCTCTGTGCAACGGGCTCAGCGGACCTGGCAACCTGAGCAGCGG  
 TCCCTACGAGGCGGGGGGAGAGAAGGAGCAGGACCCCGCTCATCCCGGTCCCTGCCAACAGCTACTTTGGCTT  
 CTCTATTGACTCGGGGAAAGGTCTGGTGGTGGTGGTGGCTGAGCTTTGTGGCTGGAGGCCCGCCGCGCAACCA  
 CAAGGGTGTGTGGTCACTCTGCGCAAGGACAGCGCCAGTGGCTGGTGGCGGAGGTATGTCTGTCTGGGAGCG  
 CTGACCTTCGGCTTTGGCTACTCACTGGCTGGCTGACCTCAACAGTGTATGGCTGGCCAGACTGATAGTGGG  
 TGCCCTCTACTTCTTTGAGCGCAAGAAAGAGCTGGGGGGTGTGTGTATGTGACTTGAACAGGGGGGTCACTG  
 GGCTGGGATCTCCCTCTCGGCTCTCGGCTCCCTGACTCATGTTCCGGATCAGCTGGCTGTCTGGGGGA  
 CCTCAACCAAGATGGCTTTCCAGATAATTCAGTGGGTGGCTCCCTTTGATGGTATGGGAAAGTCTTCATCTACCA  
 TGGGAGAGCGCTGGGCTGTGCGCAAACTCTCAAGGTGCTGGAGGGCGAGGCTGTGGGCTCAAGAGCTTCGG  
 CTACTCCCTGTCAAGCAGCTTGGATGATGGATGGGAAACCAATACCTGACCTGTGGTGGGCTCCCTGGCTGACAC  
 CGCAGTGTCTTTCAGGCGCAGACCCATCCTCCATGTCTCCCATGAGGTCTTATGTCTCACAGAGCATGACCT  
 GGAGAGCCCACTGTGTCTGGCGCCACTCGGTCTGTGTGGACCTTAAGGGTCTGTTTTCAGCTACATTTCGACTCC  
 CAGCAGCTATAGGCTTACTGTGGCCCTGGAATATGTGTAGATGGGACACAGACCGGAGGCTTCGCGGGCCAGGT  
 TCCCGTGTGAGCTTCTTGAGCCGTAACTTGGAAAGAACCCAGCAGCAGGCTCCGGGCAAGGTGTGGCTGAAGCA  
 CAGACATGACCGAGTCTGTGGAGAGCCCATGTTCAGCTCCAGGAAATGTCAAGAGCAAGCTTCCGGCCATTGT  
 AGTGACCTTGTCTACAGTCTCCAGACCCCTCGGCTCGGCGACAGGCTCTTGGCCAGGGGCTGCTCCAGTGGC  
 CCCATCTCTCAATGGCCACACGCGCAGCAGCCAGCGGGCAGAGACTCACTTCTCAAGCAAGGCTGTGGTGAAGA  
 CAAGATGCTCCAGAGACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT  
 CTCGCCATGGATGTGGATGGAAACAAACGCTGTTCGACTGAGTGGAGCCAGCATTTGGCTTGGCTGAGTGTAT  
 GGTCAACCACTGCCATCGACCGAGCCAGCCAGGCTGATGGGATGATGCCATGAGGCTGAGGCTTCCGCTGT  
 CATGCTCTCTGACTCACTGCACTACTCAGGGGTCGGGGCTTGGACCTTGGAGAGGCACTCTGCTCTCCAA  
 TGAGAAATGCTCCCATGTGAGTGTGAGCTGGGGAACCCATGAGAGAGGTGCCAGTCACTTCTATCTCAT  
 CCTTAGCACTCTCGGGATCAGCTTTAGAGCAACGGAATCGAGGTGAGAGCTGTGTGGCCAGATTCAGTGAACA  
 GGAGCTGCATCCAGTCTCGACCGAGCCGCTGTCTTCTATGAGCTGCCACTGTCCATTCAGAGGCTTCCGCTTCC  
 CAGCACTCTCTTCTCTGTGTGGTGGGCGGAGAGAGCTGCACTCTGAGCGGATGTGGGCGAGCAGGT  
 CAAGTATGAGGTCAAGGTTTCCAAACCAAGGCCAGTCCGTCAGAACCTTGGGCTTCTGCTTCACTCATGTG  
 GCTCATGAGATTGCCAATGGGAAGT  
 GCAAGAAAGGCTTGTCTCTCCAGGCCCAACATCTCCAGCTGGATGTGGACAGTATAGGAGTAGGAGGGCGCGGA  
 GCTGGAGCCACTGAGCAGCAGGAGCTGTGTGAGCGGCGAGGAGCCAGCATGTCTGTGGTGGCTGCTCTTGTG  
 TGAAGAAAGAAAGAACTCACTCGACTGCGCCCGGGGCGAGGCCAATGTGTGGTGTGTGTGTGTGTGTGTGTGT  
 CAGCTTTGACCGCGCGCTGTCTGCTGTCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCT  
 TGTGAACTCTCTGGAGAGT  
 TGTCTTCAAGT  
 CATCTCTCTGT  
 CAAAGCGGCGAAGCACCCGAGGCGCACGTGTGCCAGTACCATGTGGTGAAGATTCTTCGGGAAGACCGGACAGCA  
 GTTCAGAGGGAGAGAGAGCGGACCATCTGAGGAGAACATGTGGGCGAGCCCGCGGGAGGGGCGGATGACCA  
 CCCATCTCTGGCTGTGCTGACAGGCTCCGAGCTGGGCGCCGATGGGCTCCAGGCGAGGACCGCTTATGTTTCC  
 CATGTGCCAGCTGGCCTGTGGTCTGCCCTCCATCCCTTCTCCAGAGATGGCTCTTGTGGATTCAGAGGGGTAGAT  
 GGGCTGT  
 TCTTCCACCAACTTCCCTTATAGT  
 TGAGAGGGGAGGGGTGTCTGATGCAAAAGTGGGAGAGGGGCTTAATCTCTCTCTCCATTCACACTGTG  
 GTAAACAGGACCCCAAGGACGCTGCTCCCGGGAAGTGCCCTAACTAGAGGGTGGGAGGAGGTGTGTGTGTGT  
 CTAGGGCTGCTCTCTCTAGTTTCCCTCTCATCTGACCTTAGTTTGTGTGCTCAGTGTAGTGGTTTGTGTGT  
 TTGCTATTATTAAAAAATATTGAGAACAAAAAAGAAAAAAGAAAAA

1007681-102401

## FIGURE 177

></usr/seqdb2/sst/DNA/Dnaseqs.min/ss.DNA55737

><subunit 1 of 1, 1141 aa, 1 stop

><MW: 124671, pI: 5.82, NX(S/T): 5

MAGARSRDPWGASGICYLFGSLLVELLFSRAVAFNLDVMGALRKEBGEFGSLFGFSVALHRQL  
QPRPQSWLLVGAPQALALPGQANRTGGLFACPLSLEETDCYRVDDIQGADMQKESKENQWL  
GVSVRQQGGKKIVTCAHRYEARQRVDQILETRDMIGRCFVLSQDLAIRDELDDGGEWKFCG  
RPQGHEQFGFCQQGTAAAFSPDSHYLLFGAPGTYNWKGTFARVELCAQGSADLAHLDDGPYEA  
GGEKEQDPRLIPVPANSYFGFSIDSGKGLVRAEELSFVAGAPRANHKGAVVILRKDSASRLV  
PEVMLSGERLITSGFGYS LAVADLNSDGWPD LIVGAPYFFERQBELGGAVVYVYLNQGGHWAGI  
SPLRLCGSPDSMFGISLAVLGD LNQDGF PDIAVGAPFDGDGKVFTIYHGSSSLGVVAKPSQVLE  
GEAVGIKSPGYSLSGSLDMDGNQYPDLLVGS LADTAVLFRAPILHVSHEVSIAPRSIDLEQ  
PNCAGGHSVCVDL RVCF SYIAVPSSYSPTVALDYVL DADTDRRLRGQVPRVTFLSRNLEEPK  
HQASGTVWLKHQHDRVCGDAMFQLQENVKDKLRAIVVTLSYSLQTPRLRRQAPQGGLPPVAP  
ILNAHQPTQRAEIHFLKQCGGEDKICQSNLQLVHARFCTRVSDTEFQPLPMDVDGTTALFA  
LSGQPVIGLELMVTNLPSDPAQPQADGDDAHEAQLLVMLPDSLHYSGVRALDPAEKPLCLSN  
ENASHVECELG NPMKRG AQVTFYLILSTSGIS IETTELEVELLATISEQELHPVSARARVF  
IELPLSITAGMAIPQQLFFSGVVRGERAMQSERDVGSKVKYEVTVSNQGS LRTLGS AFLNIM  
WPHEIANGKWLLYPMQVELEGGQGPQKGLCSFRPNILHLDVDSRDRRRRELEPPEQQEPGE  
RQEPMSMWPPVSSAEKKKNITLDCARGTANCVVFSCLPYSFDRAAVLHVWGRLLWNSTFLEEY  
SAVKSL EIVVRANTVTKSSIKNLMRLDASTVIPVMVYLDPM AVVAEGVPWWVILLAVLAGLL  
VLALLVLLLVKMGVFFKRAKHPEATVPQYHAVKIPREDRQQFKBEKTGTILRNNGWSPRREGP  
DAHPILAADGHPGLGPDGHPGPGTA

### Important features:

#### Signal peptide:

amino acids 1-33

#### Transmembrane domain:

amino acids 1040-1062

#### N-glycosylation sites.

amino acids 86-89, 746-749, 949-952, 985-988 and 1005-1008

#### Integrins alpha chain proteins.

amino acids 1064-1071, 384-408, 1041-1071, 317-346, 443-465, 385-407, 215-224, 634-647, 85-99, 322-346, 470-479, 442-466, 379-408 and 1031-1047

**FIGURE 178**

CCGCGCCGGGCGCAGGAGAGCTAGTGGACGGCTCTGAGACGGCGCGCGTGCAGCAGCTCCAGA  
AAGCAGCGAGTTGGCAGAGCAGGGCTGCATTTCAGCAGGAGCTGCAGACACAGTGCTGGCT  
CACAAACAAGATGCTCAAGGTGTGAGCCGTACTGTGTGTGTGTGCAGCCGCTTGGTGCAGTCA  
GTCTCTCGCAGCTGCCCGCGGCGGTGGCTGCAGCCGGGGGGCGGTGCGAGCGCGGTAATTTTC  
TGGATGATAAACAATGGCTCACCAATCTCTCAGTATGACAAGGAAGTCGGACAGTGGAAAC  
AAATTCGAGACGAAGTAGAGGATGATTATTTCCGCACTTGGAGTCCAGGAAAACCCCTTCGA  
TCAGGCTTTAGATCCAGCTAAGGATCCATGCTTAAAGATGAAATGTAGTCGCCATAAAGTAT  
GCATTGCTCAAGATTCTCAGACTGCAGTCTGCATTAGTCACCGGAGGCTTACACACAGGATG  
AAAGAAGCAGGAGTAGACCATAGGCAGTGGAGGGGTCCCATATTATCCACCTGCAAGCAGTG  
CCCAGTGGTCTATCCAGCCCTGTTTGTGGTTTCAGATGGTCATACCTACTCTTTTCAGTGC  
AACTAGAATATCAGGCATGTGTCTTAGGAAAAACAGATCTCAGTCAAATGTGAAGGACATTGC  
CGATGTCTCTCAGATAAGCCCACCAGTACAAGCAGAAATGTTAAGAGAGCATGCAGTGACCT  
GAGGTTTCAGGGAAGTGGCAAAACAGATTGCGGGACTGGTTCAAGGCCCTTCATGAAAGTGGAA  
GTCAAAACAGAAAGACAAAACATTTGCTGAGGCTGAGAGAACGAGCATTCGATACCAGCATC  
TTGCCAATTGTGCAAGGACTCATTGGCTGAGTGTTAACAGACTGATACAAACTGTAGCTT  
GCTATTGGACCAGTGCAGAGCTCAGAAGCATTTACCTTGATAAGAATGAACAGTGTACCAAGG  
CATTCCTCAATTCTGTGACACATACAAGGACAGTTTAAATCTAATAATGAGTGGTGCTAC  
TGCTTCCAGAGACAGCAAGACCCACCTTGCCAGACTGAGCTGCAGCAATATTTCAGAAGCGGCA  
AGGGGTAAGAAGCTCCTAGGACAGTATATCCCCTGTGTGATGAAGATGGTTACTACAAGC  
CAACACAATGTCTGCGCAGTGTTGGACAGTGTGGTGTGTTGACAGATATGGAATGAAGTC  
ATGGGATCAGAATAAATGGTGTGTCAGATTGTGCTATAGATTTTGAGATCTCCGGAGATTT  
TGCTAGTGGCGATTTTCATGAATGGACTGATGATGAGGATGATGAAGACGATATTATGAATG  
ATGAAGATGAAATTGAAGATGATGATGAAGATGAAGGGGATGATGATGATGGTGGTGTGAC  
CATGATGTATACATTGATGATGACAGTTGAAATCAATAAAATTTCTACATTTCTAATATTTA  
CAAAAATGATAGCCTATTTAAAAATTATCTTCTTCCCAATAACAAAATGATTCTAAACCTCA  
CATATATTTTGTATAATTATTTGAAAAATTGCAGCTAAAGTTATAGAACTTTATGTTTAAAT  
AGAATCATTGTCTTTGAGTTTTTATATTCTTACACAAAAAGAAAAATACATATGCAGTCTA  
CTGAGACAAAATAAGTTTTGAAGTGCTACTATAATAAATTTTTTCACGAGAACAAACTTTGT  
AAATCTCTCAATAAGCAAAATGACAGCTAGTGTCTGGGATCGTACATGTTAATTTTTTGAAAG  
ATAATTCTAAGTGAAATTTAAATAAATAAATTTTTTATGACCTGGGCTCTTAAGGATTTAGG  
AAAAATATGCATGCTTTAATTGCATTTTCAAAGTAGCATCTCTGTAGACCTAGATGAGTCAG  
GATAACAGAGAGATACCACATGACTCCAAAAA

## **FIGURE 179**

></usr/seqdb2/sst/DNA/Dnaseqs.min/ss.DNA49829  
><subunit 1 of 1, 436 aa, 1 stop  
><MW: 49429, pI: 4.80, NX(S/T): 0  
MLKVSAVLCVCAAAWCSQSLAAAAVAAAGGRSDGGNFLDDKQWLTTISQYDKEVGQWNKFR  
DEVEDDYFRTWSPGKPFQDALDPAKDPCLMKCSRHKVCIAQDSQTAVCISHRRRLTHRMKEA  
GVDHRQWRGPILSTCKQCQPVVYPSFVCGSDGHTYSFQCKLEYQACVLGKQISVKCEGHCCPCP  
SDKPTSTSRNVKRACSDLEFREVANRLRDWFKALHESGSQNKTKTLLRPERSRFDTSILPI  
CKDSLGMFNRDLTDNYDLLLDQSELRSIYLDKNEQCTKAFFNSCDTYKDSLISNNEWCYCFQ  
RQQDPPCQTELSNIQKRQGVKKLLGQYIPLCDEDGYKPTQCHGSVGCWCVDRYGNEVMGS  
RINGVADCAIDFEISGDFASGDFHEWTDDEDEDDIMNDEDEIEDDEDEGDDDDGGDDHDVYI

**Important features:**

**Signal peptide:**

amino acids 1-16

**Leucine zipper pattern.**

amino acids 246-267

**N-myristoylation sites.**

amino acids 357-362, 371-376 and 376-381

**Thyroglobulin type-1 repeat proteins**

amino acids 353-365 and 339-352

# FIGURE 180

CAGACTCCAGATTTCCTGTCAACCACGAGGAGTCCAGAGAGGAAACGCGGAGCGGAGACACAACAGTACCTGACGC  
 CTCTTTTCAGCCCGGGATTCGCCACGACGAGGATGGCGGACAGAATCTGGCTTGCCTTCCCGTGCTCTTCTGGCC  
 GCTCTGCCCTCGGTGCTGCTGCTGGGGCGGCGCGGCTTTCACACTTCCCTCGATAGCGACTTCACTTTTACCGTT  
 CCGCGCGGCGAGAAGGAGTGTCTTACCAGCCATGCCCTGAAGGCCCTCGCTGGAGATCGAGTACCAAGTTTATA  
 GATGGAGCAGGATTAGATATTGATTTCCATCTTGCCTCTCCAGAAGGCCAAACCTTAGTTTGTGACAAAGAAA  
 TCAGATTGAGATTCAACATCTAGAGACTGAAGTGGTGATTACATGTTCTGCTTTGACAAATACATTACAGCAATTT  
 TCTGCGAAGGTGATTTTCTTTGAATTAACTCTGGATAAATGGGAGAACAGGCAAGAAACAAGAGATTGGAG  
 AATATATATCTGCGACAGATATATTGGATATGAACTGGAAGACATCCTGGAAATCCATCAACAGCATCAAGTCC  
 AGACTAAGCAAAAGTGGGCACATACAAATCTGCTTAGAGCATTTGAAGCTCGTGATCGAAACATACAGAGAAAGC  
 AACCTTTGATAGAGTCAATTTCTGGTCTATGGTTAATTTAGTGGTCACTGGTGGTGGTGTGACGCATTCAAGTTTAT  
 ATGCTGAAGAGTCTGTGTTGAAGATAAGAGGAAAAGTAGAACTTAAACTCCAAACTAGAGTACGTAAACATTGAAA  
 AATGAGGCATAAAAAATGCAATAAATGTTACAGTCAAGACCATTAAATGGTCTTCTCCAAAAATATTTGAGATATA  
 AAAGTAGGAAACAGGTATAATTTTAATGTGAAAAATTAAGTCTTCACTTCTGTGCAAGTAACTCTGCTGATCCAG  
 TTGTACTTAAGTGTGTAACAGGAATATTTGCGAATAATAGGTTTAACTGAATGAAGCCATTAATAACTGCAAT  
 TTTCTAACTTTGAAAAATTTTGCAAAATGTCTTAGGTGATTTAAATAAATGAGTATTGGGCGCTAATTGCAACACC  
 AGTCTGTTTAAACAGGTTCTATTACCCAGAACTTTTGTAAATGGCGGAGTTACAAATTAACCTGGGAAGTTT  
 TCGACTTTTAAGTTATAAACTACCTGAGAAATACCTAATGATGGATTGAATAAATCTTTAGACTACAAAGAGCCAA  
 CTTTTCTCTATTATACATATGATCTCTCCATAATGTAAATAGAATAATAGCTTTGAAATACAAATTTGGTCTTTTG  
 AGATTTTATAACCAAAATACATTTTCAGTGTAAACATATAGCAGAAAGCATTTAGTCTTTGTATCTTTAGCTTACATCT  
 CCAAAGCTGACATTTTTCAGGATTTTAAAAACACAAAGTTTACACTTACTAAATTTAGGACATGTTTCTCTTTG  
 AAATGAAGAAATATAGTTTAAAGCTTCTCTCCATAGGGACACATTTCTCAACCTTAACATAAAGTGTAGGA  
 TTTTAAATTAATGTGAGGTAAATAAGTTTATTTTAAATAGTATCTGTCAAGTTAATATCTGTCAACAGTTAA  
 TAATCATGTATGTAAATTTTAAACATGATTGCTGACTTGGATAATTCATTTATACCGAGTATTGAAGGAAATA  
 TTGCTAAATGATCTGGGCTAACCAATAAATAATCTCTTTCTGAGCTCTAAGATTTATCAAGAAACAGGAA  
 AGAATTTAGAAAACTTGAGAAACCTTAATCCAAATAAATAATTCATTAAGTAGAATTTAATAAATATATCAGA  
 TCTGACTGGCTCATCATGACATCTTACTCAATAACATAAATCAAAGGATGATTAATTTCCAGTTAGCTGGAG  
 AAATTTGGCTGTAGTTTATTTTCTACAGAATTTCTGGTTTGAATTTTGTGAAGCAGGTACATTTTATA  
 AAAATGGGCCCTATCTGTAAGTTTTCAGCTGGGTGTACATATTTTATAAAAAATTTTATTAACAACTTTTAT  
 TAAAAATGGCTTTCTGACACCTTTATTTTATGATGTTGAAGTAAGGATAGAACTAGACTCCCAAGTTTAAAA  
 CACCTAAATGTGAATAAACCAATATATACAAAGGTTTCTGCGATCTAGCTTTTGAAGCTATAGGGGGTCTTAC  
 TCAAGTACTAGTAAATTTAACTTCACTCATGAATGAATCTTATTTGAGCTATAGGCGGTCTTAC  
 GACTACATTTGATGATTAGAAACAACTTAAATTTGGGGTATAGAACCCCTCAACAGGTTAGTAATGCTGGAATT  
 CTGTAGACGATAATGATATGATACAGAGAGTGATTTCATTTACACTCATAGTAGTATAAAAGAGATACATTTCCC  
 TCTTAGGCCCTCGGAGAGAGACAGCTTAGATTTCCCTACTGGCAAGGTTTAAAAAATGAGTAAATGCCGTAT  
 ATGACTAATTAACCTTAAATGCCACAGAAAAATGCTTCAGGTGTCTAGGGGTATCTCTGCAACACTTGACAGAACAA  
 AGGTCAATAAGACTTCTGCTCATGAATAACCCCTCCCTTTTGGCGTGTAAATTTGCAATGAGAAGCAAAATTTACA  
 GTACCAATCAATAAAGCAGGGTCACAGATATAAACTACTGCTCTTTCTATAAAATCTGTATTAAGAATTTCTA  
 CTTCTCTGTATGGCTTACTGTACTGTACTCTCTGACTCTTCACTAACCAAGTAATTTGATATAAATCTTCT  
 ACATGTATGATTTTGGCCACTGATCTTAAACCTATGATCAGTAACCTTCTACCAATATAAAACGATAATTTGCTT  
 TATTTGAAAAAGAAATTTAGGAATCACTAAGGACATTAATTTTATAGACAAAGTAAAAAGACAGATATTTAAGAGG  
 CATACCAAAAAAGCAAAATCTGTAAACAGAGTAAAAATCTTAAATATTTCTAAGACATATCTGTTTATCTGCTT  
 CATATGCTTTTAAATTTCACTATTCATTTCTAAATTAAGTATGCTAAATTTAGTAAAGCTGTTTATCACTT  
 AACAGCTCATTTTGTCTTTTCAATATACAAAATTTTAAAAATACTACAATATTTAACTAAGGCCCAACCGGATTT  
 CATATAAGTACGATTACCGTGTCACTCACTCAAGGCTAGAGTTTGTCTGATATGCAATTTGATGATTAAT  
 GTTATGCTGTTCTTCTCAGTGAATGTCAAGCATGGAGGGTGTGTTAAATTTATGTGAAAAATTAATCCCTCTTA  
 CACATAATGGTGTCTTAAAAATGACAAAAATGAGCACCTACAATGTATGTCTCTCAATGAAGATTTCTTAT  
 GTGAAATTTTAAAGACATGATTTCCGATGTAAAGATTTTCACTGTAAGTACAAATATGCACAATCAGTGTG  
 CTCAAATCGCTTTACTACTTATAAACAGCCATCTCTAAATAGCAACATTTGTGAGTACTGATATGATATAATAA

10017081.102401

## FIGURE 181

></usr/seqdb2/sst/DNA/Dnaseqs.min/ss.DNA52196  
><subunit 1 of 1, 229 aa, 1 stop  
><MW: 26017, pI: 4.73, NX(S/T): 0  
MGDKIWLPPFVLLLAALPPVLLPGAAGFTPSLSDSFTFTLTPAGQKECFYQPMPLKASLEIEY  
QVLDGAGLDIDFHLASPEGKTLVFEQRKSDGVHTVETEVGDYMFCDNTFTSTISEKVIFTEL  
ILDNMGEQAQEQEDWKYITGTDILDMKLEDILESINSIKSRLSKSGHIQILLRAFEARDN  
IQESNFDVRVNFWSMVNLVVMVVSAIQVYMLKSLFEDKRKRSRT

### **Important features:**

#### **Signal peptide:**

amino acids 1-23

#### **Transmembrane domain:**

amino acids 195-217

#### **N-myristoylation site.**

amino acids 43-48

#### **Tyrosine kinase phosphorylation site.**

amino acids 55-62

10017351-102401

## **FIGURE 182**

CCATCCCTGAGATCTTTTTATAAAAAACCCAGTCTTTGCTGACCAGACAAAGCATACCAGAT  
CTCACCAGAGAGTGCAGACACTATGCTGCCTCCCATGGCCCTGCCAGTGTCTCTGGATG  
CTGCTTTCTGCCTCATTCTCCTGTGTCAGGTTCAAGGTGAAGAAACCCAGAAGGAACTGCC  
CTCTCCACGATCAGCTGTCCCAAAGGCTCCAAGGCCTATGGCTCCCCCTGCTATGCCTTGT  
TTTTGTCACCAAATCCTGGATGGATGCAGATCTGGCTTGCCAGAAGCGGCCCTCTGAAAA  
CTGGTGTCTGTGCTCAGTGGGGCTGAGGGATCCTTCGTGTCCTCCTGGTGAGGAGCATTAG  
TAACAGCTACTCATACATCTGGATTGGGCTCCATGACCCACACAGGGCTCTGAGCCTGATG  
GAGATGGATGGGAGTGGAGTAGCACTGATGTGATGAATTACTTTGTCATGGGAGAAAAATCCC  
TCCACCATCTTAAACCTGGCCA CTGTGGGAGCCTGTCAAGAAGCACAGGATTTCTGAAGTG  
GAAAGATTATAACTGTGATGCAAAGTTACCCTATGTCTGCAAGTTCAAGGACTAGGGCAGGT  
GGGAAGTCAGCAGCCTCAGCTTGGCGTGCAGCTCATCATGGACATGAGACCAGTGTGAAGAC  
TCACCTTGAAGAGAATATTCTCCCCAACTGCCCTACCTGACTACCTTGTGATGATCCTCC  
TTCCTTTTCTTTTCTTACCTTCATTTCAAGCTTTTCTGTCTTCCATGTCTTGAGATC  
TCAGAGAATAATAATAAAAAATGTTACTTTATAAAAAAAAAAAAAAAAAAAAAA

## FIGURE 183

</usr/seqdb2/sst/DNA/Dnaseqs.min/ss.DNA56965  
<subunit 1 of 1, 175 aa, 1 stop  
<MW: 19330, pI: 7.25, NX(S/T): 1  
MLPPMALPSVSWMLLSCLILLCQVQGEETQKELPSPRISCPKGSKAYGSPCYALFLSPKSWM  
DADLACQKRPSGKLVSVLSGAEGSFVSSLVRSISNSYSYIWIGLHDPTQGSEPDGDGWEWSS  
TDVMNYFAWEKNPSTILNPGHCGSLSRSTGFLKWKDYNCDAKLPYVCKFKD

### **Important features:**

#### **Signal peptide:**

amino acids 1-26

#### **C-type lectin domain signature.**

amino acids 146-171

100708102400

## **FIGURE 184**

CCAGTCTGTGCGCCACCTCACTTGGTGTCTGCTGTCCCCGCCAGGCAAGCCTGGGGTGAGAGC  
ACAGAGGAGTGGGCCGGGACCAATGCGGGGACGCGGCTGGCGCTCCTGGCGCTGGTGTGGC  
TGCTTGGGAGAGCTGGCGCGGCCCTGCGCTGCTACGTCTGTCCGGAGCCCAAGGAGTGT  
CGGACTGTGTCACCATCGCCACCTGCACCACCAACGAAACCATGTGCAAGACCACACTCTAC  
TCCCGGGAGATAGTGTACCCCTTCCAGGGGGACTCCACGGTGACCAAGTCTGTGCCAGCAA  
GTGTAAGCCCTCGGATGTGGATGGCATCGGCCAGACCCCTGCCCGTGTCTGCTGCAATACTG  
AGCTGTGCAATGTAGACGGGGCGCCCGCTCTGAACAGCCTCCACTGCGGGGCCCTCACGCTC  
CTCCCACTCTTGAGCCTCCGACTGTAGAGTCCCCGCCACCCCATGGCCCTATGCGGCCCA  
GCCCCGAATGCCTTGAAGAAGTGCCCCCTGCACCAGGAAAAAAAAAAAAAAAAAAAA

## **FIGURE 185**

</usr/seqdb2/sst/DNA/Dnaseqs.min/ss.DNA56405

<subunit 1 of 1, 125 aa, 1 stop

<MW: 13115, pI: 5.90, NX(S/T): 1

MRGTRLALLALVLAACGELAPALRCYVCPEPTGVSDCVTIATCTTNETMCKITLYSREIVYP  
FQGDSTVTKSCASKCKPSDVGIGQTLFVSCCNTELCNVDGAPALNSLHCGALTLLPLLSRL

**Important features:**

**Signal peptide:**

amino acids 1-17

**N-glycosylation site.**

amino acids 46-49

1007331-102404

## FIGURE 186

CTGCAGTCAAGGACTCTGGGACCGCAGGGGGCTCCCGGACCCTGACTCTGCAGCCGAACCGGC  
ACGGTTTCGTGGGGACCCAGGCTTGCAAAGTGACGGTCATTTTCTCTTTCTCCCTCTT  
GAGTCCTTCTGAGATGATGGCTCTGGGCGCAGCGGGAGCTACCCGGGTCTTTGTGCGGATGG  
TAGCGGCGGTCTCGGCGGCCACCCCTCTGCTGGGAGTGAGCGCCACCTTGAAGTCGGTTCTC  
AATTCCAACGCTATCAAGAACCTGCCCCACCGCTGGGCGGCGCTGCGGGGCACCCAGGCTC  
TGCAGTCAGCGCCGCGCCGGGAATCCTGTACCCGGGCGGGAATAAGTACCAGACCAATTGACA  
ACTACCAGCCGTACCCGTGCGCAGAGGACGAGGAGTGCGGCACTGATGAGTACTCGCTAGT  
CCCACCCGCGGAGGGGACGCGAGCGTGCAAATCTGTCTCGCCTGCAGGAAGCGCCGAAAACG  
CTGCATGCGTCACGCTATGTGCTGCCCCGGGAATTACTGCAAAAATGGAATATGTGTGTCTT  
CTGATCAAAATCATTTCCGAGGAGAAATTGAGGAAACCATCACTGAAAGCTTTGGTAATGAT  
CATAGCACCTTGGATGGGTATTCCAGAAGAACCACCTTGTCTTCAAAAATGTATCACACCAA  
AGGACAAGAAGGTTCTGTTTGTCTCCGGTCATCAGACTGTGCCTCAGGATTGTGTGTGCTA  
GACACTTCTGGTCCAAGATCTGTAAACCTGTCTTGAAGAAGGTCAAGTGTGTACCAAGCAT  
AGGAGAAAAGGCTCTCATGGAAGTAAATAATCCAGCGTTGTTACTGTGGAGAAGGCTGTGC  
TTGCCGGATACAGAAAGATCAACCATCAAGCCAGTAATTTCTTAGGCTTCACACTTGTGAGA  
GACACTAAACCAGCTATCCAAATGCAGTGAACCTCTTTTATATAATAGATGCTATGAAAACC  
TTTTATGACCTTCATCACTCAATCCTAAGGATATACAAGTTCTGTGGTTTCAGTTAAGCAT  
TCCAATAACACCTTCCAAAACCTGGAGTGTAAGAGCTTTGTTTCTTTATGGAACCTCCCTG  
TGATTGCAGTAAATTAAGTATTGTAAATTTCTCAGTGTGGCACTTACCTGTAAATGCAATGA  
AACTTTTAATTAATTTTCTAAAGTGCTGCACTGCCTATTTTCTCTTGTATGTAAATTT  
TTGTACACATTGATGTTATCTTGACTGACAAATATTCTATATTGAACTGAAGTAAATCATT  
TCAGCTTATAGTTCTTAAAGCATAACCCCTTACCCCATTTAATTCTAGAGTCTAGAAGCGA  
AGGATCTCTTGGAAATGACAAATGATAGGTACCTAAAATGTAACATGAAAATACTAGCTTATT  
TTCTGAAATGACTATCTTAATGCTTAAATTATATTTCCCTTTAGGCTGTGATAGTTTTTGA  
AATAAAATTTAACATTTAAAAAAAAAAAAA

## **FIGURE 187**

</usr/seqdb2/sst/DNA/Dnaseqs.min/ss.DNA57530

<subunit 1 of 1, 266 aa, 1 stop

<MW: 28672, pI: 8.85, NX(S/T): 1

MMALGAAGATRVFVAMVAAALGGHPLLGVSATLNSVLNSNAIKNLPPPLGGAAGHPGSAVSA  
APGILYPGGNKYQTIDNYQPYPCAEDDEECGTDEYCASPTRGGDAGVQICLACRKRKRRCMRH  
AMCCPGNYCKNGICVSSDQNHFRGEIIEETITESFGNDHSTLDGYSRRITLSSKMYHTKGQEG  
SVCLRSSDCASGLCCARHFWSKICKPVLKEGQVCTKHRRKGSHGLEIFQRCYCGEGLScriQ  
KDHHQASNSSRLHTCQRH

**Important features:**

**Signal peptide:**

amino acids 1-23

**N-glycosylation site.**

amino acids 256-259

**Fungal Zn(2)-Cys(6) binuclear cluster domain**

amino acids 110-126

## FIGURE 188

TGTGTTTCCCTGCAGTCAGAAATTTGGGACNGCAGGGGTTCCCGGACCTGATTTGCAGCGGA  
ACGGGAAGGTTTTGTGGGACCCAGGTTGAAATGACGGTCATTTTTTTTCTTCTCCTTCNG  
GAGTCCTTNTGAGANGATGGTTTTTGGGCGCAGCGGGAGCTAACCCGGTTTTTTGTNGCGATG  
GTAGCGGCGGTTTTTCGGCGGCCACCTINTGCTGGGAGTGAGCGCCACCTTGAATCGGTTTTC  
AATTCGAACGNTATCAAGAACCTGCCCCACCGNTGGGCGGCGCTGCGGGGCACCCAGGNTT  
TGCAGTCAGCGCCGCGCCGGGAATCCTGTACCGGGCGGGAATAAGTACCAGACCATTGACA  
ATTACCAGCCGTACCCGTGCGCAGAGGACGAGGAGTGCGGCACTGATGAGTACTGCGCTAGT  
CCCACCCGCGGAGGGGANGCGGGCGTGCAAATNTGTNTNGCCTGCAGGAAGCGCCGAAAACG  
CTGCATGCGTCANGCTATGTGCTGCCCCGGGAATTACTGCAAAAATGGAATATGTGTGTNTT  
CTGATCAAAATCATTTCCGAGGAGAAATTGAGGAAACCATCACTGAAAGCTTTGGTAATGAT  
CATAGCACCTTGGATGGG

10017081-102401

# FIGURE 189

GAGGAACTACCGGTACCGGCGCGCGCTGGTAGTCGCGGTTGGCTGCACTCACCAATCCCGTGGCGCGCGG  
CTGGGCGGTGGAGAGTGGCTGTGCTTCTCTCTGCAAGCGGTGCTTGGGCTCGGCGAGGCGGGTCCGCGGCA  
GGGTTTGAGGATGGGGGATGACTACAGGAAGCGACCCCGGATGGCAAGGTATTTTTTGTGGAATGAAAAGAA  
AGTATTAGAAATGAGCTGAAGACCAATCACAGATTAAATTTTTTGGGACAGATTTTTGTGATGCTGTAATCCACCT  
TGAGTAAATGTAGACAGAAAGTTCTCAAAATTTGCATATTACATCAACTGGAACAGCAGTGAATCTTAATGTTTCA  
TTAAATCAGAACTTGCATAAGAAAGAAATGCGAGTCTGGTTAAATAAGATGACATATATAGAGAGTGAAGAG  
GATCATTCTCTGTTTTCTGATAGTGTATATGGCCATTTAGTGGGCACAGATCAGGAATTTTACAGTTTACTTTGG  
AGTGTCCTAAATCTGCAGACAGATAGAAAATTAAGACAAGCTTTCAAGAAATTTGGCATTTGAAGTTACATCCGTGATAA  
AARCCCGAATAACCCAAATGCATAGGCGATTTTTTAAATAAATAGAGCATATGAAGTACTCAAGAGTGAAGA  
CTATCGGAAAAAGTATGACAAATATGGAGAAAAGGGAATCTGAGGATTAATCAAGGTGCGCAATATGAAGAGTGAAGA  
TTCTATTCTGTTATGATTTTTGTGATTTATGATGATGATCCTGAAATCATAAACATGGAAGAAGAGAAATTTGATGTC  
TGCTGTTAAATCTGGAGAACTGTGGTTTTGTAAATTTTTACTCCCCAGGCTGTTTCACTGCCATGATTAGCTCC  
CACATGGAGAGACTTTTGTAAAGAAAGTGGATGGGTACTTCGAATTTGGAGCTGTTAACTGTGGTGATGATAGAAT  
GCTTTGGCGAATGAAAGGAGTCAACAGCTATCCAGCTCTCTCATTTTTTCGGTCTGGAATGGCCCCAGTGAAATA  
TCATGGACAGATCAAGAGGAGGTTTGTAGGTTTTGCAATGCAGCATGTTAGAAGTACAGTGACAGAACTTTG  
GACAGGAAATTTTGTCACTCCATCAAACTGCTTTTGTCTGCTGGTATTTGGCTGGCTGATCACTTTTTGTTCAA  
AGGAGGAGATTTTGTACTTCACAGACAAGCTCAGGCTAGTGGCATGTTGTTTCTCAACTCATTTGGATGCTAA  
AGAAATATATTTGGAAGTAATACATAATCTCCAGATTTTGAACACTCTTTGCGGAAACACCTAGAGGATCGTT  
GGCTCATCATCGTGGCTGTTTTTCTATTTTGAATAAATGAAATCAATGATCTCGAGCTGAAAAAAT  
AAAACTCTCACTTAAAAATGATCATATCAAGTTGGCAGGTTTGAATGTTCTCTGACAGACATCTGTAGTAA  
TCTGTATGTTTTTCAAGCGTCTCTTCAGCATTTTAAAGGACAAGGACCAAGAAATATCATCATGGAAGA  
GAAGATTCTATGATATATCTGCTTTGCAAAAGAAAGTGTGAATTTCTCATGTTTACCAAGCTGTGGACCTCAAAA  
TTTTCTGCGCAATGCAAAAGAACCATGGCTTGTGATTTCTTTGCCCCCTGGTGTCCACCATGTGCGAGCTTTACT  
ACCAGAGTACGAAGAGCATCAATCTCTTTATGGTCACTTAAGTTTGGTATCACTAGATTTGTACAGTTTCATGA  
GGAGCTCTGTAACCTTCAAGCAATTTCAAGCTTTATCCAAACAAGTGGTATTCACCAAGCTTCAAGTATCATGAGTA  
TGAAGGACATCATCTGCTGAGCAAAATCTTGGAGTTTCAAGAGGATCTTATGAATCTCTCAGTGGTCTCTCTTAC  
ACCCACCTTCTCAAGCAATTTGATTTACAAAGAAAAACAACGAAGTCTGATGGTTGTTTCTATTCTCTCGTG  
GTGTCATCTTGGCAAGCTTTAATGCGCAAGTGGAAAGAAATGGCCCGGACATTAATCTGATGATCAAGTGGG  
CAGTATGATTTGGCAACAGTATCATCTTTTGTGGCCAGGAAACGTTCAAGATACCTGAGTAAAGATTTTTT  
TCCCCCAAAATCAAAATAAGCTTTATCAGTATCAAGTACAATGCTTGAATAGGGATGCTTATCTCCCTGAGAA  
CTGGGCTTAGGATTTTAACTCAAGTATCCAAGATCTAACCACTCAGACTTTCACTGAGTAAAGATTTCTACAG  
GAAAAATCATTTGGGTGATTTTCTATCTCTCTGCTGGAACCTTGGCAGAAATTTTGGCTGAGTATTTAGCT  
CTTGGCTAGGATGATTAAGAGAAAGTGAAGCTGGAAAGTAGACTGTGAGGCTTATGTCATGACATGCCAGAA  
AGCTGGATCAGGCGCTATCCAACCTGTTAAGTTTTATTTCTACGAAAGAGCAAGAGAAATTTCAAGAAAGCA  
GATAAATACCAAGATGCAAAAGCAATCGCTGCTTAAATAGTGAAAAATTTGGAACCTCTCCGAAATCAAGGCAA  
GAGGAATAAGGATGAATTTGATAAGTTTGAAGATGAAGAAAAAGTTTAAAGAAATTTCTGACATGATCATG  
AAGACACCTATTAGAAATGTACATTTATGATGGGAATGAATGAACATTATCTAGACTTGCAGTTGTATGTCGCA  
GAATATTCTACAGCACTGGTGTAAAGAGGGTCTGCAAACTTTTCTGTAAGGCGGGTTTATAAATATTTTA  
GACTTTGCGAGGCTATAATATATGTTTCAACATGAGAACAGAAATAGAGTTCATCATGATATTTCTGTTATTTGCT  
TTTTAACCACTTTTAAAAAATATTAACAGATTCTTAGCTCAGAGCATAAAAGTAGGCTGGATTCAGTCCATG  
GACCATAGATTGCTGCTCCCTCGACGGAGTTATAATGTTTCAAGTGGCTGGCTGAACTGAGTCTGCTGTGCT  
ATCTACATAAGTGTCTAAGTTGTATAAAGTCCATTTTCCCTTCAAGTTTTTGGCTGACCTGAAAGAGGTAAT  
TAGTTTTTGGTCACTGTTTCTCTTAAATGCTATCCCTAACCATATATTTATATTTTCTGTTTTTAAAAACCCAT  
GATGTGGCAAGTAAACCAACCTTTATGCTGTATTATATGAGGAGATTTCTCATTTGTTTCTTTCTTCTTCA  
AAGTTTGAAGAAATCTTTTAAATTTTTCACAGCCGAGAAACAGTGCAGCAGTATATGTGCAACAGTAAAGTAC  
AAATTTGAAACAGTAAAGTGCACAAATTTCTGTATTTTGTGTATCATCCAGGAAACCTGAGGGAACAAATTA  
TAGCAATTAACCTGGGCACTTGAAGATATCCTAAATATGTTATCAAGTATTTAGAGTCTATATTTTAAAGATATA  
TGTTTTCTCATGATTTTCTGAATTTGCTTTTATAGAAATTTTCCCATGATGATTTGATTTTGGGCACTGAATAT  
TTACATATCTGCTCTGAACTTTGTTTTGACCTGTATCCTTTATTTACATTTGGTTTTTCTTCTCATAGTTTTGG  
TTTTTCACTCTGCTCAGTCTATTTATTTATTAATCAAAATAGGAAAAATTTACATTCAGTTGTTTTTCTGTAGCTTAT  
AATGATCTCTAGTTTATTCAGTTTACAGTTTACTGTGAGAGGCTGCTTTTTTCTAGATAAATATGACATATA  
ACTGAAGTATTTTTTATAGAAATCAAGTATATAAATCAGGAAAGGATCTTCTGATTTCTGTTGTTTGA  
CTCAAGAAATCACAATTTGTGAGTAAATGATGTTGTTAGTTATATAATTCAGAGTGTACAGAAATGTTAAAT  
CCAATCAGTCAAGAGAGTCAATGAATTAAGGCTTGCACCTTTTTTCAAAAAAAAAAAAAAA

## **FIGURE 190**

</usr/seqdb2/sst/DNA/Dnaseqs.min/ss.DNA56439  
<subunit 1 of 1, 747 aa, 1 stop  
<MW: 86127, pI: 7.46, NX(S/T): 2  
MGVWLNKDDYIRD LKRIILCF L I V M A I L V G T D Q D F Y S L L G V S K T A S S R E I R Q A F K K L A L K L  
H P D K N P N N P N A H G D F L K I N R A Y E V L K D E D L R K K Y D K Y G E K G L E D N Q G G Q Y E S W N Y Y R Y D F G I  
Y D D D P E I I T L E R R E F D A A V N S G E L W F V N F Y S P G C S H C H D L A P T W R D F A K E V D G L L R I G A V N C  
G D D R M L C R M K G V N S Y P S L F I F R S G M A P V K Y H G D R S K E S L V S F A M Q H V R S T V T E L W T G N F V N S  
I Q T A F A A G I G W L I T F C S K G G D C L T S Q T R L R L S G M L F L N S L D A K E I Y L E V I H N L P D F E L L S A N  
T L E D R L A H H R W L L F F H F G K N E N S N D P E L K K L K T L L K N D H I Q V G R F D C S S A P D I C S N L Y V F Q P  
S L A V F K G Q G T K E Y E I H H G K I L Y D I L A F A K E S V N S H V T T L G P Q N F P A N D K E P W L V D F F A P W C  
P P C R A L L P E L R R A S N L L Y G Q L K F G T L D C T V H E G L C N M Y N I Q A Y P T T V V F N Q S N I H E Y E G H H S  
A E Q I L E F I E D L M N P S V V S L T P T T F N E L V T Q R K H N E V W M V D F Y S P W C H P C Q V L M P E W K R M A R T  
L T G L I N V G S I D C Q Q Y H S F C A Q E N V Q R Y P E I R F F P P K S N K A Y Q Y H S Y N G W N R D A Y S L R I W G L G  
F L P Q V S T D L T P Q T F S E K V L Q G K N H W I D F Y A P W C G P C Q N F A P E F E L L A R M I K G K V K A G K V D C  
Q A Y A Q T C Q K A G I R A Y P T V K F Y F Y E R A K R N F Q E E Q I N T R D A K A I A A L I S E K L E T L R N Q G K R N K D E L

**Important features:**

**Endoplasmic reticulum targeting sequence.**

amino acids 744-747

**Cytochrome c family heme-binding site signature.**

amino acids 158-163

**Nt-dnaJ domain signature.**

amino acids 77-96

**N-glycosylation site.**

amino acids 484-487

## FIGURE 191

AGACAGTACCTCCTCCCTAGGACTACACAAGGACTGAACCAGAAGGAAGGACAGAGCAAA  
GCCATGAACATCATCTAGAAATCCTTCTGCTTCTGATCACCATCATCTACTCTACTTGGAG  
GTCGTTGGTGAAAGTTTTTCATTCTCTCAGAGGAGAAAATCTGTGGCTGGGGAGATTGTTCTCA  
TTACTGGAGCTGGGCATGGAATAGGCAGGCAGACTACTTATGAATTTGCAAAACGACAGAGC  
ATATTGGTTCTGTGGGATATTAATAAGCGCGGTGTGGAGGAAACTGCAGCTGAGTGC CGAAA  
ACTAGGCGTCACTGCGCATGCGTATGTGGTAGACTGCAGCAACAGAGAAGAGATCTATCGCT  
CTCTAAATCAGGTGAAGAAAGAAGTGGGTGATGTAACAATCGTGGTGAATAATGCTGGGACA  
GTATATCCAGCCGATCTTCTCAGCACCAAGGATGAAGAGATTACCAAGACATTTGAGGTCAA  
CATCTAGGACATTTTTGGATCACAAAAGCACTTCTTCCATCGATGATGGAGAGAAATCATG  
GCCACATCGTCACAGTGGCTTCAGTGTGCGGCCACGAAGGGATTCCTTACCTCATCCCATAT  
TGTTCCAGCAAATTTGCCGCTGTTGGCTTTACAGAGGCTCGACATCAGAACTTCAGGCCTT  
GGGAAAACTGGTATCAAAAACCTCATGTCTCTGCCAGTTTTTGTGAATACTGGGTTCCACCA  
AAAATCCAAGCACAAAGATTATGGCCTGTATTGGAGACAGATGAAGTCGTAAGAAGTCTGATA  
GATGGAATACTTACCAATAAGAAAATGATTTTTTGTTCATCGTATATCAATATCTTCTGAG  
ACTACAGAAGTTTCTTCTGAACGCGCCTCAGCGATTTTAAATCGTATGCAGAATATTCAAT  
TTGAAGCAGTGGTTGGCCACAAAATCAAAATGAAATGAATAAAATAAGCTCCAGCCAGAGATG  
TATGCATGATAATGATATGAATAGTTTGAATCAATGCTGCAAGCTTTATTTACATTTTT  
TCAGTCTGTATAATATATAAACATTGGTTTGGCACTAGCAGCAGTCAAACGAACAAGATTA  
ATTACCTGTCTTCTGTTTTCTCAAGAATATTTACGTAGTTTTTTCATAGGTCTGTTTTTCCTT  
TCATGCCTCTTAAAAAATTCTGTGCTTACATAAACATACTTAAAAGGTTTTCTTTAAGATAT  
TTTTTTTTTCCATTTAAAGGTGGACAAAAGCTACCTCCCTAAAAGTAAATACAAAGAGAACT  
TATTTACACAGGGAAGGTTTAAAGACTGTTCAAGTAGCATTTCCAATCTGTAGCCATGCCACAG  
AATATCAACAAGAACAAGAATGAGTGCACAGCTAAGAGATCAAGTTTCAGCAGGCAGCTTT  
ATCTCAACCTGGACATATTTTAAAGATTGACATTTTGAAAGATTTCCCTAGCCTCTTCTTTTT  
TCATTAGCCCAAAACGGTGCAACTCTATTCTGGACTTTATTACTTGATTCTGTCTTCTGTAT  
AACTCTGAAGTCCACAAAAGTGACCCCTCTATATTCTCTCCCTTTTTATAGTCTTATAAGA  
TACATTATGAAAGGTGACCGACTCTATTTTAAATCTCAGAATTTTAAAGTTCTAGCCCCATGA  
TAACCTTTTTCTTTGTAATTTATGCTTTTCATATATCCCTGGTCCAGAGATGTTTAGACAAT  
TTTAGGCTCAAAAATTAAGCTAACACAGGAAAAGGAAGTGTACTGGCTATTACATAAGAA  
CAATGGACCCAAGAGAAGAA

## **FIGURE 192**

</usr/seqdb2/sst/DNA/Dnaseqs.min/ss.DNA56409

<subunit 1 of 1, 300 aa, 1 stop

<MW: 33655, pI: 9.31, NX(S/T): 1

MNIILEILLLLITIIYSYLESIVKFFIPQRRKSVAGEIVLITGAGHGIGRQTTYEFAKRQSI  
LVLWDINKRGVEETAABCRKLGVTAHAYVVDCSNREEIYRSLNQVKKEVDVTIVVNNAGTV  
YPADLLSTKDEEITKTFEVNILGHFWITKALLPSMMERNHGHIVTVASVCGHEGIPYLIPYC  
SSKFAAVGFHRLTSELQALGKTGIKTSCLCPVFVNTGFTKNPSTRLWPVLETDEVVRSIID  
GILTNNKMIFVPSYINIFLRLQKFLPERASAILNRMQNIQFEAVVGHKIKMK

**Important features:**

**Signal peptide:**

amino acids 1-19

**cAMP- and cGMP-dependent protein kinase phosphorylation site.**

amino acids 30-33 and 58-61

**Short-chain alcohol dehydrogenase family protein**

amino acids 165-202, 37-49, 112-122 and 210-219

## FIGURE 193

CGGCGGCGGCTGCGGGCGCAGGTTGAGGGGCGCGAGGTTGAGGGGCGCGAGGTTCCAGCAGG  
ATGCCCGGCTCTGCAGGAAGCTGAAGTGAGAGGCCCGAGAGGGCCAGCCCGCCGGGGC  
AGGATGACCAAGGCCCGGCTGTTCCGGCTGTGGCTGGTGTGGGGTCGGTGTTCATGATCCT  
GCTGATCATCGTGTACTGGGACAGCGCAGGCGCCGCGCACTTCTACTTGACACGTCCTTCT  
CTAGGCCGCACACGGGGCCGCGCTGCCACGCCCGGGCCGGACAGGGACAGGGAGCTCAGC  
GCCGACTCCGATGTCGACGAGTTTCTGGACAAGTTTCTCAGTGCTGGCGTGAAGCAGAGCGA  
CCTTCCAGAAAGGAGACGGAGCAGCCGCTGCGCGGGGAGCATGGAGGAGAGCGTGAGAG  
GCTACGACTGGTCCCCGCGCAGCGCCCGGCGCAGCCAGACCAGGGCCGGCAGCAGGCGGAG  
CGGAGGAGCGTGCTGCGGGGCTTCTGCGCAACTCCAGCCTGGCCTTCCCCACCAAGGAGCG  
CGCATTCGACGACATCCCCAACTCGGAGCTGAGCCACCTGATCGTGGACGACCCGGCACGGGG  
CCATCTACTGCTACGTGCCCAAGGTGGCTGCACCAACTGGAAGCGCGTGATGATCGTGCTG  
AGCGGAAGCCTGCTGCACCGGGTGCGCCTACCGCGACCCGCTGCGCATCCCGCGCGAGCA  
CGTGACACAACGCCAGCGCGCACCTGACCTTCAACAAGTTCTGGCGCGCTACGGGAAGCTCT  
CCCGCCACCTCATGAAGGTCAAGCTCAAGAAGTACACCAAGTTCTCTTCGTGCGCGACCCC  
TTCGTGCGCCTGATCTCCGCCTTCCCGCAGCAAGTTGAGCTGGAGAACGAGGAGTTCTACCG  
CAAGTTGCGCGTGCCCATGCTGCGGCTGTACGCCAACCAACCAGCCTGCCCGCCTCGGCGC  
GCGAGGCCTTCCGCGCTGGCCTCAAGGTGTCTTCCGCCAACTTCATCCAGTACCTGCTGGAC  
CCGCACACGGAGAAGCTGGCGCCCTTCAACGAGCACTGGCGGCAGGTGTACCGCCTCTGCCA  
CCCGTGCCAGATCGACTACGACTTCGTGGGGAAGCTGGAGACTCTGGACGAGGACGCCGCGC  
AGCTGCTGCAGCTACTCCAGGTGGACCGCAGCTCCGCTTCCCCCGAGCTACCGGAACAGG  
ACCGCCAGCAGCTGGGAGGAGGACTGGTTGCGCAAGATCCCCCTGGCCTGGAGGACGAGCT  
GTATAAAGCTCTACGAGGCCGACTTTGTTCTCTTCGGCTACCCCAAGCCGAAAACCTCCTCC  
GAGACTGAAGCTTTCGCGTTGCTTTTCTGCGGTGCCTGGAACTGACGACGCGCACTCC  
AGTTTTTTTATGACCTACGATTTTGCAATCTGGGCTTCTTGTTTCACTCCACTGCCTCTATCC  
ATTGAGTACTGTATCGATATTGTTTTTAAGATTAAATATATTTCAGGTATTTAATACGA

## **FIGURE 194**

</usr/seqdb2/sst/DNA/Dnaseqs.min/ss.DNA56112

<subunit 1 of 1, 414 aa, 1 stop

<MW: 48414, pI: 9.54, NX(S/T): 4

MTKARLFRLWLVLGVSFMILLIIIVYWDSAGAAHFVYLHTSFSRPHTGPPPLTPGPDRLDELTA  
DSDVDEFLDKFLSAGVKQSDLPRKETEQPPAPGSMEEVSVRGYDWSPRDARRSPDQGRQQAER  
RSVLRGFCANSSLAFFTKERAFDDIPNSELSHLIVDDRHGAIYCYVPKVACTNWKRVMIIVLS  
GSLLRGAPYRDPLRIPREHVHNASAHLTFNKFWRRYGKLSRHLMKVKLKKYTKFLFVRDPF  
VRLISAFRSKFELENEFFYRKFAVPMRLRYANHTSLPASAREAFRAGLKVSFANFIQYLLDP  
HTEKLAPFNEHWRQVYRLCHPCQIDYDFVGKLETLDDEAQLLQLLQVDRQLRFPFSYRNRT  
ASSWEEDWFAKIPLAWRQQLYKLYEADFVLFGYPKPENLLRD

### **Important features:**

#### **Signal peptide:**

amino acids 1-31

#### **N-glycosylation sites.**

amino acids 134-137, 209-212, 280-283 and 370-373

#### **TNFR/NGFR family cysteine-rich region protein**

amino acids 329-332

[illegible]

TCTGGGCCAGAATTCTGCGCACGAGGCGGCACGAGGGCGACGGCCTCACGGGGCTTTGGAGGTGA  
AAGAGGCCCCAGAGTAGAGAGAGAGAGAGACCGACGTACACGGGATAGGCTACGGGAACCGCCT  
ATGCCGGGAAGGTGTGTGCTGTGACCGGGGGCGGGCGCGGCATCGGAGCTGGGATCGTGCGC  
GCCCTTCGTGAACAGCGGGGCCCGAGTGGTTATCTGCGACAAGGATGAGTCTGGGGGCGGGC  
CCTGGAGCAGGAGCTCCCTGGAGCTGTCTTTATCCTCTGTGTAGTGTGACTCAGGAAGATGATG  
TGAAGACCTTGTTTCTGAGACCATCCGCCGATTTCGGCCGCTGGATTGTGTTGTCAACAAC  
GCTGGCCACCACCACCCCCACAGAGGCCTGAGGAGACCTCTGCCACGGGATTCCGCGCAGCT  
GCTGGAGCTGAACCTACTGGGGACGTACACCTTGACCAAGCTCGCCCTCCCTACCTGCGGA  
AGAGTCAAGGGAATGTATCAACATCTCCAGCCTGGTGGGGGCAATCGGCCAGGCCCAGGCA  
GTTCCCTATGTGGCCACCAAGGGGGCAGTAACAGCCATGACCAAAGCTTTGGCCCTGGATGA  
AAGTCCATATGTGTGTCGAGTCAACTGTATCTCCCCAGGAAACATCTGGACCCCGCTGTGGG  
AGGAGCTGGCAGCCTTAATGCCAGACCCTAGGGCCACAATCCGAGAGGGCATGCTGGCCAG  
CCACTGGGCCGCGATGGGCCAGCCCGCTGAGGTCGGGGCTGCGGCAGTGTTCTGGCCCTCCGA  
AGCCAACCTTCGACAGGGCATTGAATGCTCGTGACGGGGGTGCAGAGCTGGGGTACGGGT  
GCAAGGCCAGTCGGAGCACCCCGTGGACGCCCCCGATATCCCTTCTGATTCTCTCATTT  
CTACTTGGGCCCCCTTCTTAGGACTCTCCACCCCAAACCTCCAACCTGTATCAGATGCAGC  
CCCCAAGCCCTTAGACTCTAAGCCAGTTAGCAAGGTGCCGGGTCAACCTGCAGTTCCCAT  
AAAAACGATTTGCAGCC

## FIGURE 196

</usr/seqdb2/sst/DNA/Dnaseqs.min/ss.DNA56045

<subunit 1 of 1, 270 aa, 1 stop

<MW: 28317, pI: 6.00, NX(S/T): 1

MATGTRYAGKVVVVTGGGRGIGAGIVRAFVNSGARVVICDKDESGGRALEQEELPGAVFILCD  
VTQEDDVKTLVSETIRRFGRLLDCVVNNAGHHPPPQRPEETSAQGFRQLLELNLLGTYTLTKL  
ALPYLRKSQGNVINISLVAIGQAQAVPYVATKGAVTAMTKALALDESPYGVVNCISPEN  
IWTPLWEELAALMPDPRATIREGMLAQPLGRMGQPAEVGAAVFLASEANFCTGIELLVTTGG  
AELGYGCKASRSTPVDAPDIPS

**Important features:**

**N-glycosylation site.**

amino acids 138-141

**Short-chain alcohol dehydrogenase family protein**

amino acids 10-22, 81-91, 134-171 and 176-185

## **FIGURE 197**

AGGCGGGCAGCAGCTGCAGGCTGACCTTGACGCTTGGCGGAATGACTGGCCTCACAACTG  
CTGTTTCTTCTTACCATTTCATCTTCTGGGGCTGGGCCAGCCAGGAGCCCCAAAAGCAA  
GAGGAAGGGCAAGGGCGGCTGGGCCCTGGCCCTGGCCCTCACCAGGTGCCACTGGACC  
TGGTGTCACGGATGAAACCGTATGCCCGCATGGAGGAGTATGAGAGGAACATCGAGGAGATG  
GTGGCCCACTGAGGAACAGCTCAGAGCTGGCCAGAGAAAGTGTGAGGTCAACTTGCAGCT  
GTGGATGTCCAACAAGAGGAGCCTGTCTCCCTGGGGCTACAGCATCAACCACGACCCAGCC  
GTATCCCCGTGGACCTGCCCGAGGCACGGTGCCCTGTGTCTGGGCTGTGTGAACCCCTTCACC  
ATGCAGGAGGACCGCAGCATGGTGAGCGTGCCGGTGTTCAGCCAGGTTCTGTGCGCCGCCG  
CCTCTGCCCGCCACCGCCCCGCACAGGGCCTTGCCGCCAGCGCGCAGTCATGGAGACCATCG  
CTGTGGGCTGCACCTGCATCTTCTTGAATCACCTGGCCAGAAGCCAGGCCAGCAGCCCGAGA  
CCATCCTCCTTGACCTTTGTGCCAAGAAAGGCCTATGAAAAGTAAACACTGACTTTTGAAA  
GCAAG

## FIGURE 198

</usr/seqdb2/sst/DNA/Dnaseqs.min/ss.DNA59294

<subunit 1 of 1, 180 aa, 1 stop

<MW: 20437, pI: 9.58, NX(S/T): 1

MDWPHNLLFLLTISIFLGLGQPRSPKSKRKGGQGRPGPLAPGPHQVPLDLVSRMKPYARMEEY  
ERNIEEMVAQLRNSSELAQRKCEVNLQLWMSNKRSLSPWGYSINHDPRIIPVDLPEARCLCL  
GCVNPFTMQEDRSMVSVPVFSQVPVRRRLCPPPPRTGPCRQRAVMETIAVGCTCIF

### **Important features:**

#### **Signal peptide:**

amino acids 1-20

#### **N-glycosylation site.**

amino acids 75-78

#### **Homologous region to IL-17**

amino acids 96-180.

## FIGURE 199

GCGCCGCCAGGCGTAGGCGGGGTGGCCCTTGCCTCTCCCGCTTCTTGAACCCCGCGGG  
CGAGCGAGGCTCGGGGCCGCCGCTGCCCTTCCCCACACTCCCGCCGAGAACCTCGCTCG  
GCGCCCAACATGGCGGGTGGGCGCTGCGGCCCGCAGCTAACGGCGCTCTTGGCCGCTGGAT  
CGCGGCTGTGGCGGCGACGGCAGGCCCGAGGAGGCCGCGCTGCCCGGAGCAGAGCCGGG  
TCCAGCCCATGACCGCTCCAACCTGGACGCTGGTGATGGAGGGCGAGTGGATGCTGAAATTT  
TACGCCCCATGGTGTCCATCCTGCCAGCAGACTGATTCAGAATGGGAGGCTTTTGCAAGAA  
TGGTGAAATACCTTCAGATCAGTGTGGGGAAGGTAGATGTCATTCAAGAACAGGTTTGAGTG  
GCCGCTTCTTTGTCAACCACCTCTCCAGCATTTTTTCATGCAAAAGGATGGGATATTCGCCGT  
TATCGTGGCCAGGAATCTTCGAAGACCTGCAGAATTATATCTTAGAGAAGAAATGGCAATC  
AGTCGAGCCTCTGACTGGCTGGAATCCCCAGCTTCTCTAACGATGCTCGGAATGGCTGGTC  
TTTTTAGCATCTCTGGCAAGATATGGCATCTTCACAACATTTTCACAGTGACTCTTGGAATT  
CCTGCTTGGTGTCTTATGTGTTTTCTGTCATAGCCACCTTGGTTTTTGGCCTTTTTATGGG  
TCTGGTCTTGGTGGTAATATCAGAATGTTTCTATGTGCCACTTCCAAGGCATTTATCTGAGC  
GTTCTGAGCAGAATCGGAGATCAGAGGAGGCTCATAGAGCTGAACAGTTGCAGGATGCGGAG  
GAGGAAAAAGATGATTCAATGAAGAAGAAAAACAAGACAGCCTTGATAGATGATGAAGAAGA  
GAAAGAAGATCTTGGCGATGAGGATGAAGCAGAGGAAGAAGAGGAGGAGGACAACCTTGGCTG  
CTGGTGTGGATGAGGAGAGAAGTGAGGCCAATGATCAGGGGCCCCAGGAGAGGACGGTGTG  
ACCCGGGAGGAAGTAGAGCCTGAGGAGGCTGAAGAAGGCATCTCTGAGCAACCCCTGCCCAGC  
TGACACAGAGGTGGTGGAAGACTCCTTGAGGCAGCGTAAAGTCAGCATGCTGACAAGGGAC  
TGTAGATTTAATGATGCGTTTTCAAGAATACACACCAAAACAATATGTCAGCTTCCTTTGG  
CCTCAGATTTGTACCAAACTCTTAATTTTTCTGAATGAGCAAGCTTCTCTTAAAGATGCT  
CTCTAGTCATTTGGTCTCATGGCAGTAAGCCTCATGTATACCTAAGGAGAGCTTCCAGGTGT  
GACAATCAGGATATAGAAAAACAACGCTAGTGTGGGATCTGTTTGAGACTGGGATGGGAA  
CAAGTTCAATTTACTAGGGGTGAGAGAGTCTCGACCAGAGGAGGCCATTCCAGTCCTAATC  
AGCACCTTCCAGAGACAAGGCTGCAGGCCCTGTGAATGAAAGCCAAGCAGGAGCCTTGGCT  
CCTGAGCATCCCCAAAGTGTAACGTAGAAGCCTTGCACTCTTTTCTGTGTAAAGTATTTAT  
TTTTGTCAAATTGCGGAAACATCAGGCACCACAGTGCATGAAAACTTTTCACAGCTAGAA  
ATTGAAAGGGCCTTGGGTATAGAGAGCAGCTCAGAAGTCATCCAGCCCTCTGAATCTCCTG  
TGCTATGTTTTTACTTTACCTTTAATTTTTCCAGCATTTCCACCATTGGGCATTCAGGCTCT  
CCCACTCTTCACTATTATCTCTTGGTCAGAGGACTCCAATAACAGCCAGGTTTACATGAAC  
TGTGTTTGTTCATTTGACCTAAGGGGTTTAGATAATCAGTAACCAATACCCCTGAAGCTGT  
GACTGCCAAACATCTCAAATGAAATGTTGTGGCCATCAGAGACTCAAAGGAAGTAAGGATT  
TTACAAGACAGATTAAAAAAAATGTTTTGTCCAAAATATAGTTGTTGTGATTTTTTTTT  
AAGTTTTCTAAGCAATATTTTTCAAGCCAGAAGTCTCTAAGTCTTGCCAGTACAAGGTAGT  
CTTGTGAAGAAAAGTTGAATACTGTTTTGTTTTTCATCTCAAGGGGTTCCCTGGGTTCTGAAC  
TACTTTAATAATAACTAAAAAACCACTCTGTATTTCTTCAGTGATGTGCTTTTGGTGAAA  
GAATTAATGAACCTCAGTACTGAAAGTGAAAGATTGATTTTGTTCATCTCTGTGAATC  
TTCCAAAGAATTATATCTTTGTAAATCTCTCAATACTCAATCTACTGTAAGTACCCAGGGAG  
GCTAATTTCTTT

## **FIGURE 200**

</usr/seqdb2/sst/DNA/Dnaseqs.min/ss.DNA56433  
<subunit 1 of 1, 349 aa, 1 stop  
<MW: 38952, pI: 4.34, NX(S/T): 1  
MAGGRCGPQLTALLAAWIAVAATAGPEEAALPPEQSRVQPMTASNWTLVMEGEWMLKFYAP  
WCPSCQQTDSSEWFAFAKNGEILQISVGKVDVIQEPGLSGRFFVTTLPAFFHAKDGIFRRYRG  
PGIFEDLQNYILEKKWQSVEPLTGWKSPASLTMSGMAGLFSISGKIWHLHNYFTVTTLGIPAW  
CSYVFFVIATLVFGLFMGLVLVVISSECFYVPLPRHLSESEQNRRSBEAHRAEQLQDAEEEEK  
DDSNEEENKDSLVDDEEEKEDLGDEDEAEEDNLAAGVDEERSEANDQGPPGEDGVTRE  
EVEPEEAEEGISEQPCPADTEVVEDSLRQRKSKHADKGL

### **Important features:**

#### **Signal peptide:**

amino acids 1-22

#### **Transmembrane domain:**

amino acids 191-211

#### **N-glycosylation site.**

amino acids 46-49

**Thioredoxin family proteins.** (homologous region to disulfide isomerase)

amino acids 56-72

#### **Flavodoxin proteins**

amino acids 173-187



## **FIGURE 202**

</usr/seqdb2/sst/DNA/Dnaseqs.min/ss.DNA53912  
<subunit 1 of 1, 201 aa, 1 stop  
<MW: 22563, pI: 4.87, NX(S/T): 1  
MEYHPDLENLDEDGYTQLHFDSQSNTRIAVVSEKGS CAASPPWRLIAVILGILCLVILVIAV  
VLGTMGVLSPPCPNWIIEYKSCYLFMSLNSWDGSKRQCWQLGSNLLKIDSSNELGFIVKQ  
VSSQPDNSFWIGLSRPQTEVPWLWEDGSTFSSNLFQIRTTATQENPSPNCVWIHVSVIYDQL  
CSVPSYSICEKKFSM

### **Important features:**

#### **Type II transmembrane domain:**

amino acids 45-65

#### **cAMP- and cGMP-dependent protein kinase phosphorylation site.**

amino acids 197-200

#### **N-myristoylation sites.**

amino acids 35-40 and 151-156

#### **Homologous region to LDL receptor**

amino acids 34-67 and 70-200.

# FIGURE 203

GGAAGGGGAGGAGCAGGCCACACAGGCACAGGCCGGTGAAGGACCTGCCAGACCTGGAGGGTCTCGCTCTGTGCA  
 CACAGGCTGGAGTGCAGTGGTGTGATCTTGGCTCATGTAACCTCCACTCCGGGTCAAAGTGTCTCATCTGATGCC  
 TCAGCTCTCCCGAGTAGCTGGGATTACAGGTGGTGAATCTCCAGAGTAGTACTCCGTGGGAGGAAATAGCATCCCCAG  
 TCGCTGCTCAGACGACACTGTTCTGCTGAGTCTGCTCTTCTCGGTCCAGAGTGCCACAGGCGAGGGGCCACAGG  
 GAGACTTTCGCTTCTGCGAGCCAGCGAACCCAGACACACAGGAGCAGCTTCCATCAAAACCCACACAGACCTTG  
 CGACTCTCCATCAGAACTCCGAAGAGGGCCCTCAGAGTCCATGCCCTTTCCTCGAGCCACCTGCTTCCCGA  
 TCCTTCTCTGACCCGAGGGCTCTACCACTTCTGCGCTCTACTGGAACGACATGCTGGGAGATTACATCTTCTC  
 TATGGCAAGCGTGACTTCTTGTCTGAGTGACAAAGCCTCTAGCCTCCTCTGCTTCCAGCACCAGGAGGAGCGCTG  
 GCTCAGGGGCCCCGCTGTTAGGCATCTCTGTCACTCTGGTGGAGCCCTCAGAACATCAGGCTGCCAGTGGCC  
 GCCAGCTTCACTTCTCTCCATCAGTCTCTCCACACAGCGCCGCTCACAATGCTCGGTGGACATGTGCGAGCTC  
 AAAAGGGACCTCAGCTGCTCAGCGAGTCTCTGAAGCATCCCCAGAAGGCCCTCAAGGAGGCCCTCGGCTGCCGCC  
 GCCAGCCAGCAGTTGTCAGAGCCTGGAGTGGAACTGACCTCTGTGAGATTCAATGGGGGACATGGTGTCTCTCGAG  
 GAGGACCCGGATCAACGCCACGGTGTGGAAGCTCCAGCCACAGCCGGCTCCAGGACCTGCACATCCACTCCGGG  
 CAGGAGGAGGAGCAGAGCGAGATCATGGAGTACTCGGTGCTGCTGCCCTCGAACACTCTTCCAGAGGACGAAGGCG  
 CGGAGCGGGGAGGCTGAGAAGAGACTCTCTCGGTGGATTCAGCAGCCAAAGCCCTGTCCAGGACAAAGATTC  
 AGCCAACTCTGGGTGAGAAGGCTCTTGGGGATTGTGGTACAGAACACCAAAGTCCAACTCAGCGAGCCCGTGT  
 GTGCTCACTTTCAGCAGCCAGCTCAGCCGAAGAATGTGACTCTGCAATGTGTGTTGAGGCTGAAGACCCCA  
 TTAGAGCAGCCCGGGGAGCTTGGAGCAGTGTCTGGGTGTGAGACCGTCAGGAGAGAAACCCAAACACTCTGCTTCTGCG  
 AACACTTGAACATACTTTGACGCTGTATGGTCTCTCGGTGGAGGTGGAAGCCCTGTGCAAGACATCACTGAGC  
 CTCTCTCTCTACGTGGGCTGTGTGCTCTCTGCCCTGGCCCTGCTTGTGACCATGCGCCCTACTCTCTGCTCAGG  
 GTGCCCGGCGTGTGAGGAGAAACCTCGGAGTACACCATCAAGGTGACATGAACGCTGCTGCGGCCGTCTTCT  
 CTGCTGGACACGAGCTTCTGCTCAGCGAGCCGGTGGCCCTGACAGGCTCTGAGGCTGGCTGCCGAGCCAGTGGC  
 ATCTTCTGCACTTCTCCCTGCTCACCTCGCTTCTCTGGATGGGCTCGAGGGGTACAACCTTACCACTGCTGTG  
 GTGGAGGTCTTTGGCAGCTTGTCTCCCTGGCTACTACTCAAGCTGAGCGCATGGCTGGGCTTCCCATCTTTT  
 CTGGTGAGCGCTGCTGCGCCCTGGTGGATGTGGACAACTATGGCCCATCATCTTGGCTGTGCATAGGACTCCAGAG  
 GGGCTCATCTACCTTCTCATGTCTGTGGATCGGGACCTCCTCGGTGAGTACATCAACCACTGGGCCCTCTTCAGC  
 CTGGTGTTCGTGTTTCCATAGGCGCATGTAGCCACCATGGTGGTGCAGATCTCGGCTGGCGCTGCGCCCCACCCAA  
 AAGTGGTGCATGTGCTGATCACTGCTGGCCCTCAGCCTGCTGCTTGTGCGCTGGCCCTGTGATCTTCTCTCC  
 TTGCTTCTGAGCACTCTCCAGCTTGTGCTGCTCTCTACTCTTTTCAGCATCATCACTCTCTTCCAGGCTTCTCATC  
 TTCTCTGCTACTGCTCCATGCGCGCTCAGGCGCCGGGTGGCCCTCCCTCTGAAAGAGCACTCAGACAGCGCC  
 AGGCTCCCATCACTCAGCTCGGCGAGCCTCTGCTCAGCCGATCATAGGCTCCAGCCACCTTGCCTATGTATGAG  
 CAGAGATGCGGCTCTGTCGACACTGCTGTGGCCCTGGAGCGAGGCCAGCCAGCCAGCCAGTCCAGCCGAGACT  
 TTGGAAGGCCCAACGACCATGAGAGATGGCGGCTTGGCATGGTGGAGCGGACTCCCGGGCTGGGCTTTTGAATGT  
 GCCTTGGGAGCACTCTGGCTCTCCTCAGCTCCACGGGACTCAGAAGTGCGCCCATGCTGCTCTAGGGAATG  
 TCCCCACATCTGTCCCCAACCCAGCTGGAGGCTGGTCTCTCTTACAAACCCCTGGGCCAGCCCTCATTTGCTGGG  
 GGGCAGGCTCTGGATCTTGAGGGTCTGGGCATCTTAACTCTGTGCCCTGCTGGGACAGAAGTGTGGCTCCA  
 GTTGTCTGTCTCTGTTGTGCTCAGCTGAGGGCACTCTGCATCTCTGTGATTTAACTCAGGTGGCACCCAGGG  
 CGAATGGGGCCCGAGGCGAGCTTTCAGGGCCAGAGCCTGGCGGAGGAGAGCCCTTTCAGGAGCACAGCAGC  
 AGCTCGCTACTCTGAGCCAGGCCCCCTCCCTCCCTCAGCCCCCAGTCTCCTCCATCTTCTCTGGGGTTC  
 TCTCTCTCCAGGGGCTCTGTCTCTCTTGGTTACAGCTGGGGTCCCCGATTCCAATGCTGTGTTTTGGGGA  
 GTGGTTTCCAGAGGCTGCTGGTGTCTGCTGTAAATGTTTGTCTACTGCAACAGCTCTGGGCTGCCCTGAGGCCA  
 GGCTCGTACCGATGCTGGGCTGGGCTAGGTCTCTCTGTCCATCTGGGCCTTTGTATGAGCTGCAATGTGCCCTTG  
 CTCACTCTGACCAACGACACAGCCTCAGAGGGGCCCTCAGCCTCTCCTGAAGCCCTCTTGTGGCAAGACTGTGGA  
 CCATGCGAGTCCCGTCTGGTTTCCATCCACCACTCCAAGGACTGAGACTGACCTCTCTGTTGACCTTGGCCTA  
 GAGCCTGACATCTCTCTAAGAGTCTCTCCTCAAGCCCCAAATAGCTCCAGGCGCCCTCGGCGGCCCATGATGT  
 TAAATCTGTCCAAACAAACACACAGGTAGATTGTGGCTGTGTGATGGGACAGAGTACAGTACGACCTG  
 GTCACTCTCTCTGCCAATATTAGTCTGTGATGTGAGGCGTGCCTGAAGCAGAACTCTCTGGAGCTCAGGGGACA  
 GGGAGCCATCACTTCTCTGGGAATCTTGAAGACTCTCTGAGGAGTCAAGTCTTCAATCTGACCTTGAAGT  
 GTTGGAGGATGTTCTTTTACGTACCAATCTTTTGTCTTTGATATAAAGAAAGTACATGTCTCATTTGTAGAGA  
 ATTTGGAACTGTAGAAGAGATCAAGAAGAAAATAAATAATCAGCTGTTGTAATGCTAGCAAAAATAAAAA  
 AA

10017031-102403

## **FIGURE 204**

</usr/seqdb2/sst/DNA/Dnaseqs.min/ss.DNA50921  
<subunit 1 of 1, 693 aa, 1 stop  
<MW: 77738, pI: 8.87, NX(S/T): 7  
MTPQSLQLQTTLFLLSLLFLVQGAHGRGHREDFRCSQRNQTHRSSLHYKPTPDLRISIENSE  
EALT VHAPFPAHPASRSFPDPRGLYHFCLYWNRHAGRLHLLYGKRDFLLSDKASSLLCFQH  
QEESLAQGPPLLATSVTSWWSPQNI SLPSAASFTFSFHSPPHTAAHNASVDMCELKRDQLL  
SQFLKHPQKASRRPSAAPASQQLSLESKLTSVRFMGDMVSFEEDRINATVWKLQPTAGLQD  
LHIHSRQEEEQSEIMEYSVLLPRTL FQRTKGRSGEAEKRLLLVDFSSQALFQDKNSSQVLGE  
KVLGIVVQNTKANLTPVVLTFQHQLQPKNVTLQCVFWVEDPTLSSPGHWSSAGCETVRR  
TQTSFCFNLHTYFAVLMVSSVEVDVAVHKHYLSLLSYGCVVVSALACLVTIAAYLCSRVP LPC  
RRKPRDYTIKVHMNLLLA VFLLDTSFLLSEPVALTGSEAGCRSAIFLHFSLLTCLSWMGLE  
GYNLYRLVVEVFQTYVPGYLLKLSAMGWGFPIFLVTLVALVDVDNYGP IILAVHRTPEGVIY  
PSMCWIRDSLVS YITNLGLFSLVFLFNMA MLATMVVQILRLRPHTQKWSHVL TLLGLSLVLG  
LPWALIFFSFASGT FQLVVLYLFSIITSFQGF LIFIWYWSMRLQARGGPSPLKSN SDSARLP  
ISSGSTSSSRI

### **Important features:**

#### **Signal peptide:**

amino acids 1-25

#### **Putative transmembrane domains:**

amino acids 382-398, 402-420, 445-468, 473-491, 519-537, 568-590  
and 634-657

#### **Microbodies C-terminal targeting signal.**

amino acids 691-693

#### **cAMP- and cGMP-dependent protein kinase phosphorylation sites.**

amino acids 198-201 and 370-373

#### **N-glycosylation sites.**

amino acids 39-42, 148-151, 171-174, 234-237, 303-306, 324-327  
and 341-344

#### **G-protein coupled receptors family 2 proteins**

amino acids 475-504

[illegible]

TGCCTGGCCTGCCTTTGTCAACAATGCCGCTTACTCTGCTTCCAGGTTGCCCTGCCTTGCGAGA  
GGAAANCNTCGGGACTACACNNTCAAGTGCACATGAACCTGCTGCTGGCCGTCTTCTGCTG  
GACACGAGCTTCTGCTCAGCGNAGCCGTTGGCCCTGACAGGCTCTGAAGGCTGGCTGCCGA  
GCCAGTGGCATCTTCTGCACTTCTCCTGCTCACCTGCCTTTCTTGATGGGCCTCGAGGGG  
TACAACCTCTACCGACTCGTGGTGAGGCTCTTTGGCACCTATGTCCCTGGCTACCTACTCAA  
GCTGAGCGCCATGGGCTGGGGCTTCCCCATCTTTCTGGTGACGCTGGTGGCCCTGGTGGATG  
TGGACAATCTGGCCCCATCATCTTGGCTGTGCATAGGACTCCAGAGGGCGTCATCTACCCCT  
TGGATGTGCTGGATCCGGGACTCCCTGGTCAGCTACATCACCACCTGGGCCCTCTTCAGCCT  
CGCTTTTCTGTTCACATGG



## FIGURE 207

MSLFGLLLTSALAGQRQGTQAESNLSSKFQFSSNKEQNGVQDPQHERIITVSTNGSIHSPR  
FPHTYPRNTVLVWRLVAVEENVWIQLTFDERFGLEDPEDDICKYDFVEVEEPSDGTILGRWC  
GSGTVPGKQISKGNQIRIRFVSDEYFPSEPGFCIHYNIVMPQFTEAVSPSVLPPSALPLDLL  
NNAITAFSTLEDLIRYLEPERWQLDLEDLYRPTWQLLGKAFVFGKRSRVVDLNLLEEVRLY  
SCTPRNFSVSIREELKRTDTIFWPGCLLVKRCGGNCACCLHNCNECQCVPSKVTKKYHEVLQ  
LRPKTGVRGLHKSLTDVALEHHEECDVCVRGSTGG

**Signal sequence:**

amino acids 1-14

bioRxiv preprint doi: <https://doi.org/10.1101/000000>; this version posted January 1, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

# FIGURE 208

CCCATCTCAAGCTGATCTTGGCACTCTCATGCTCTGCTCTCTTCAACGACACCTCTACATTCATTTTGAAGA  
 AGACTAAAAATGGTGTTCCTAATGTGGACACTGAAGAGACAAATCTTATCTCTTTTAACTAATCCTTAATTTCC  
 AAATCCTCTTGGGGCTAGATGGTTTCTTAAACACTCTGCCCTGTGATGTCACTCTGATGTCTTCCAAGAACCTATGG  
 ATCTGTGGACTGACAGACAAGCATTTGACAGAAATCTCTGGAGTATTCAGAACACAGCACTCAACCTC  
 ACCATTAAACACATACAGACATCTCCAGCGTCTCTTACAGACTGGACCATCTGTAGAGATCGATTTACAGAC  
 TGCACACTGTGTAACCTATTCCTCTGGGGTCAAAAAACACATGTGCATCAAGAGCGTGCAGATTAAACCCAGAGAC  
 TTTAGTGGACTCACTTATTTAAATCCCTTTACCTGGATGGAAACAGCTACTAGAGATCCGAGGGCTCTCCG  
 CCTAGCTTACAGCTTCTGAGCTTTGAGGCAACACATCTTTCCATCAGAAAGAGAAATCAACAGAACTGGCC  
 AACATAGAAATACCTACTCTGGGCCAAACTGTTATTATCGAAATCTTGTATTGTTTCAATTTCAATAGAGAA  
 GATGCCCTCTTAATCTTGACAAATGTTAAAGTGTCTCTCCCTGAAAGATAACATGTCAAGCGCTCCCTACTGTT  
 TTGCCATCTTCTTTTACAGAACTATATCTCTACAAACACATGATTGCAAAATCCAGAAAGATGATTTTAAATAC  
 CTCAACCAATTAACAATCTTGGACTAAGTGGAAATGGCCCTCGTTGTTATAATGCCCATTTTCTTGTGCGCG  
 TGTAAAAATAATTTCTCCCTACAGATCCCTGTAAATGCTTTTGTATGCGCTGACAGAAATAAAGTTTTCAGTCTA  
 CACAGTAACCTCTCTCAGCATGTGCCCCCAAGATGGTTTAAAGACATCAACAACTCCAGGAACCTGGATCTGTCC  
 CAAAACCTCTTGGCCAAAGAAATGGGGATGCTAAATTTCTGCAATTTCTCCCGAGCTCATCCAATTGGATCTG  
 TCTTTCAATTTTGAACCTCAGGTCTATCGTGATCTATGAATCTATCAACAGCATTTTCTTCACTGAAAGCGCTG  
 AAAATCTCGGGATCAGAGGATATGTTCTTAAAGAGTTGAAAGCTTAACTCTCGCCATTAATAATCTTCAA  
 AATCTTGAAGTTCTTGATCTTGGCACTAATTTATAAAATTTGCTAACCTCAGCATGTTTAAACATTTAAAGA  
 CTGAAAGTCATAGATCTTTCAGTGAATAAAATATCACTTCAGGAGATTCAGTGAAGTTGGCTCTGCTCAAAT  
 GCCAGAACTCTGTAGAAAGTTATGAACCCAGGCTCTGGAAACATTAACATTTTACAGATATGATAGATAGCA  
 AGGAGTTGCAGATTCAAAAACAAAGAGGCTTCTTTCATGTCTGTTAATGAAAGCTGTACAAGTATGGGCAGACC  
 TTGGATCTAAGTAAAAATAGTATATTTTGTCAAGTCTCTGATTTTCAAGCATCTTCTTCTCCTCAATGCCGT  
 AATCTGTGAGGAATCTCATTAGCCAAACTCTTAATGGCAGTGAATCCAACTTTAGCAGAGCTGAGATATTTG  
 GACTTCTCCAACACCGGCTTGATTTACTCCATTCAACAGCATTTTGAAGAGCTTCAACAACTGGAAGTTCTGGAT  
 ATAAGCAGTAATAGCCATTATTTCAATCAGAAGGAATATCTCATGTCTTAACTTACCAAGAACTTAAAGGTT  
 CTGAGAACTGATGATGAACGACATGACATCTCTTCTCCACAGCAGGACATGGAGAGTGAGTCTTCTAGA  
 ACTCTGGAATTCAGAGGAATCACTTAGATGTTTATGAGAGAGAGGTGATACAGATCTTACAACTTCTCAAG  
 AATCTGCTAAAAATAGAGGAATAGACATCTCTAAAAATTCCTAAGTTTCTTGGCTTCTGAGATTTTGTGGGT  
 ATGCTCTCAAAATTAAGAAATCTCTCTTGGCCAAAAATGGCTCAATCTTTCAGTGTGAAGAACTCAAGTGT  
 CTAAGAACCTGGAACCTTTGGACCTCAGCCACAACTGACCACTGTCCCTGAGAGATATCAACTGTCTCC  
 AGAAGCTCAAGAAATGTATCTCTTAAGAATAATCAATCAGGAGTCTGAGAGATTTTCTCAAGAGATGCTTCTC  
 CAGTTGCGATATCTGATCTCAGCTCAAAATAATCCAGATGATCCAAAGACAGCTTCCAGAAATGTCTCTC  
 AACATCTGAAGATGTTGCTTTTTCATATAATCGGTTCTTGATGACAGCAAGTCACTTCTTGGGATGTGTGATATTAACAT  
 GACATCAAGACCCAGCTGTGACCGAGTGGGTTTGGCTGAGCTGGTGGCCAACTGGAAGACCCAGAGAGAAAA  
 CATTTTAAATTTATGTTCTGAGGAAAGGGAGTGTACCGGGCAGCCAGTCTGGAAGAACTTTTCCAGAGCATAT  
 CAGCTTAGCAAAAAAGACAGTGTGTTGTGATGACAGACAAGTATGCAAGACTGAAATTTTAAAGATGCAATTTTAC  
 TTGTCCCATCAGAGGCTCATGGATGAAAAAGTTGATGTGATTATCTTGATATTTCTTGAAGAGCCCTTCAAGAG  
 TCCAAAGTTCTCTCAGCTCCGCGAAAGGCTCTGTGGGAGTTCTGTCTTGTAGTGGCCACAAAGCCGCAAGCTCAC  
 CCTACTCTGCGGAGTGTCTAAAGAACGCGCTTGGCCACAGACAATCATGTGGCCCTATAGTCAGGTGTTCAAGGAA  
 ACGGCTTAGCGCTTCTTCAAAAAACAACTGCTAGTTTACCAAGAGAGGCGCTGGC

10017081.102411

## FIGURE 209

MVFPMTLKRQILILEFNIIISKLLGARWFPKTLPCDVTLDVPKNHVIVDCTDKHLTEIPGG  
IPTNTTNLTLTINHDPDISPASFHRDLHLEIDFRCNCVPIPLGSKNNMCIKRLQIKPRSFS  
GLTYLKSLYLDGNQLLEIPQGLPPLQLLSLEANNIFSIRKENLTLANIEILYLGQNCYYR  
NPCYVSYSIEKDAFLNLTCLKVLSLKDNNVTAVPTVLPSTLTELYLYNNMIAKIQEDDFNNL  
NQLQILDLSGNCPRCYNAPFPCAPCKNNSPLOIPVNAFDALTELVRLHSHNSLQHVPPRWF  
KNINKLQELDLSQNFALAKEIGDAFLHFLPSLIQLDLSNFELQVYRASMNLSQAFSSLSKL  
KILRIRGYVFKELKSFNLSPLHNLQNLEVLDTGNTFIKIANLSMFKQFKRLKVIDLSVNKIS  
PSGDSSEVGFCSNARTSVESYEPQVLEQLHYFRYDKYARSCRFKNKEASFMSVNESCYKYGO  
TLDLSKNSIFFVKSSDFQHLISFLKCLNLSGNLISQTLNGSEFQPLAELRYLDFSNNRLDLLH  
STAFEELHKLEVLDISSNSHYFQSEGITHMLNFTKNLKVQLKLMNDNDISSSTSRTESES  
LRTLEFRGNHLDVLWREGDNRYLQLFKNLLKLEELDISKNSLSFLPSGVFDGMPPNLKNLSL  
AKNGLKSFSWKKLQCLKNLETLDLSHNQLTTVPERLSNCSRSCLKNLILKNNQIRSLTKYFLO  
DAFQLRYLDLSSNKIQMIQKTSFPENVLNNLKMILLHHNRFLCTCDAVWFVWVNHTVETIP  
YLATDVTCVGPGAHKGQSVISLDLYTCELDLTNLILFSLSSISVSLFLMVMMTASHLYFWDVW  
YIYHFCKAKIKGYQRLISPDCCYDAFIVYDTKDPATTEWVLAEVLAKLEDPREKHFNLCLEE  
RDWLPGQPVLENLSQSIQLSKKTVFVMTDKYAKTENFKIAFYLSHQRLMDEKVDVILIFLE  
KPFQKSKFLQLRKRLCGSSVLEWPTNPQAHFYFWQCLKNALATDNHVAYSQVFKETV

**Signal sequence:**

amino acids 1-26

**Transmembrane domain:**

amino acids 840-860

# FIGURE 210

GGGTACCATTCTGCGCTGCTGCAAGTTACGGAATGAAAAATTAGAACCAACAGAAAAATGGAAAAACATGTTCCCTTC  
 AGTCGTCAATGCTGACCTGTCATTTCTGCTAATATCTGGTTCTGAGGTTATGCGGCCAGGAAGAAATTTTCTA  
 GAAGCTATCCTTGATGATGAGAAAAAGCAAATGATCTAGTCTAGTCTATGCGGCAATCGTCGATACAGGAAG  
 TTCCCAAACCGGTGGGCAAAATATGTCAGACAACTAGACCTCTGTGATATTTTCATCACACATAGCAATGAAT  
 CATTTCAAGGGCTGCAAAATCTCACTAAAAATAATCTAAACCAACCCCAATGTCAGCAGCAAGGAGCGAAATC  
 CCGGATATCAAAATCAATGGCTTGAATATCAGACAGCGGGCATTCTCAACCTTAAAAAACCCTAAGGAGGTACTGCT  
 TTGAGAACCAACGAGTTTCCCCAAATACCTCTGGTTTGCCAGAGTCTTTGACAGAACTTAGTCTATTTCAAACA  
 ATATATATCAACATACTAAAGAGGGGCAATTCAGAGCTTATATAAATCTGAAAAATCTCTATTGGGCTGGAACTGCT  
 ATTTTAAACAAAGTTTGGGAGAAAAATTAACATAGAAGATGGAGTATTTGAAACCGCTGACAAATTTGGAGTTGCTAT  
 CACTATCTTTCAATCTCTTTTACACGTCGCCAACCACTGCGCAAGCTCCCTACGCAAACTTTTCTGAGCAACA  
 CCGAGATCAAAATACATTTAGTGAAGAAGATTTCAAGGAGTTGATAAATTTAACTATCTAGATTAAAGCGGGAAT  
 GTCCGAGGTGCTTCAATGCCCACTTTCCATGCGTGCCTTGATGAGTGGTGGTCTCAATTAAATAGATCGTTTGT  
 CTTTTCAAACTTGACCCAACCTTCGATACCTAAACCTCTAGCACTCCCTCAGGAAGATTAAATGCTGCGCTGGT  
 TTAATAATATGCTCATCTGAAGGTGCTGGATCTTGAATCAACTATTTAGTGGGAGAAATAGTCTCTGGGGCAT  
 TTTTAAACGATGCTGCCCGCTTAGAAATACTTGACTGTCTTTTAACTATATAAAGGGGAGTTATCCACAGCATA  
 TTAATTTTCCAGAAATCTCTAAACCTTTTGTCTCTAGCGGCATTGCAATTAAGAGGTTGGTATTTTCCAGGAAC  
 TCAGAGAGATGATTTTCAGCGCCTGATGACAGTCTCAAACTTATGACATCAACTTTGATATTAATTTTATTA  
 AGCAAAATCGATTTCAAACCTTTTCAAATTTCTCAATCTGGAATTTTACTTGTGAGAAACAGAAATCATCAC  
 ATCTTGGTAAAGAGTACCCGGCAGAGTATGCAAAATAGTCTCTTTTCAAGCTCATATCCGGAACAGACGCTCAA  
 CAGATTTGAGTTTGCACCAATCTGAACTTTTATCATTTTCAACCGTCTTTTAAAGCGCAAAATGCTGCTGCT  
 ATGGAAGAGCTTGAATTTAAGCTCTCAAGCATTTTCTTCAATGGGGCAAACTTTGAAATCTTCTCGTACA  
 TGGCTGTTTGAATCTGTCTGCAAAATAGCAATGCTCAAGTGTAAAGTGAACGAAATTTTCAAGCAATCTCTCATG  
 TCAAAATTTTGAATTTGCAAAACAAATAGACTAGACTTTGATAATGCTAGTGTCTTACTAGATGCTCGCATGG  
 AAGTTCTAGATCTCAGCTATAATTTTCACTATTTTCAAGATAGCAGGCGTAAACATCATCTAGAAATTTATTCAAA  
 ATTTTCAAAATCTTAAAGTTTAAACCTTGAGCCACCAACCAATTTATCTTTTCAACAGATAAGTATACTCTGAAA  
 GCAAGTCCCTGTAGAAATTAGTTTTCAGTGCAATCGCTTGACATTTTGTGAGATATGATGACCAACAGCTATA  
 TCTCCATTTTCAAAGTCTCAAGAACTGACACGCTGGATTTATCCCTTAAAGATGAGCAACATCCCAATG  
 AAGCATCCCTTAAATTTGCGACGAGTCTCACTGAATCATATAAATGATATATGTTAAAGTTTTTTAACTGGA  
 CATCTACAGCAGTCTTCTGCTCGAGTTGCTTGAAGTCTGAGTCTGAGGAAACAAATCTCTTTTAACTGATAGCC  
 TATCTGACTTTTACATCTTCCCTTCGACACTGCTGCTGAGCTATAAGCAATTTCCCACTTACCTCTGCTCTTC  
 TTTCTGAAGTCAGTACTGAGACCTCGATTTAAAGTTCAATCTGCTTAAAGCAATCAACAAATCCGCACTTG  
 AAACCTAAGACACCAACCAAAATTTATGTTGGAACTACACGGAAACCCCTTGAATGCACTGTGACATTTGGAG  
 ATTTCCGAAGATGGATGGATGAACATCTGAATGTCAAAATCCCGAGCTGGTAGATGTCATTTGTGCGAGTCTGT  
 GGGATCAAGAGGGGAAGGATATGTTGAGTCTGGAGCTAACAACTTTGTTTTCAGATGTCACTGCAAGTATATAT  
 TTTTCTCAGGTTCTTTATCACCACTGATGTTATGTTGGCTGCCCTGGCTACCAATTTGTTTTCATCGGAGTGT  
 GGTATATATATATGTTGTTTGTAGCTAAGGTAAAGGCTACAGGCTCTTTTCCCACTCCCAACTTTTCTATGATG  
 CTTACATTTCTTATGACACCAAGAGTCTCTGTTTACTGACTGGGTGATTAATGAGTCTGCGCTACCACTTTGAAG  
 AGAGCCGAGACAAAAAGCTTCTCTTTGCTAGAGGAGAGGGGATTGGGACCCGGGATTTGGCACTCATCGCAAC  
 TCATCGAGACATCAACCAAAGCAAGAAAAAGTATTTGTTTAAACCAAAAAATGATGCAAAAAAGCTGGAACTTTA  
 AAACAGCTTTTACTTTGGCTTTGCGAGGCTAAGGATGAGAACATGGATGTGATTTATCTCTGCTGAGGC  
 CAGTGTTCAGCATCTCAGTATTTGAGGCTACGGCAGCGGATCTGAAGAGCTCCATCTCCAGTGGGCTGACAC  
 ACCCGAAGGCGAAGGCTTGTTTTGGCAAACTCTGAGAAATGTGGTCTTGAAGTAAAGTATTACATCGATATAAC  
 ATATGATGTGCAATCCATTAAGCAATACTAACTGACGTTAAGTCATGATTTTCGCGCCATATAAAGATGCAAG  
 GAATGACATTTCTGATTTAGTTTCTATGCTATGTAAACAAATATCCCAAACCTTAGTGGTTTAAACCAACACA  
 TTTGCTGGCCACAGTTTTTGGGGTCAGGAGTCCAGGCCAGCATAACTGGGTCCTCTGCTCAGGTTGCTCAG  
 AGGCTGCAATGTAGGTTTCAACAGACATAGGCTACCTGGGCTCAGCTCATGTTGGTTGTCTTCTGATTTCA  
 ATTTCTCTGGGCTATGGCCAAAGGCTATACTCATGTAAGCCATGCGAGCTCTCCCAAGGCGAGCTTGCTTC  
 ATCAGAGCTGCAAAAGAGAGGTTGCTAGCAAGATGAAGTCACAACTTTGTAATGCAATCAAAAGAGTAT  
 ATCTCATCATCTTTGGCCATTTCTTATTTGTTAGAAGTAAACCAAGCTCCCAAGCTCCATGGGAGTGACCAAC  
 TCAGTCAGGGAAGAACGCTGAAGACCAAGATGGTGAAGCTCTGATGCTTCAAGTTGGTCTCACTCAATTTTCCCT  
 TGACTGCTGTCTGGGATGGCTGCTATCTTGATGATAGATTTGTAATATGAGGAGCGAGGATCATCTGTTGAAC  
 ATCTAGCATGCTTTACAAATTTGCTGTAACATTTTCAATCTAAGAACTTTTGGCACTGATTAAGTGTGCTTAT  
 TTAAGCTTGTGTTTATATTTATCATATCTATGGCTACATGTTTATATGCTGTGGTGGCTCGGTTTAT  
 TTAGCTTGTCTTTTACAAATTTGCTGTAACATTTTGAATTTTGAAGTATGAGTGCCATTTAAGATCTAGATGG  
 ATAGCTTTTAAAGCATCTTTTACTCTTCACTTTTAAAGATATGACGCTAAATTCGAAGCTTTTGGTCTATA  
 TTGTTAATGGCATTTGCTGTAATCTTAAATGAATGAATAAAATGTTTCACTTTTACAAAAA

1007081-102402

## FIGURE 211

MENMFLQSSMLTCIFLLISGSCELCAEENFSRSYPCDEKKQNSVIAECSNRRLQEVPTVG  
KYVTELDLSDNFITHITNESFQGLQNLTKINLNHNPNVQHONGNPGIQSNGLNITDGAFLNL  
KNLRELLLEDNQLPQIPSGLPESLTELSLIQNNIYNITKEGISRLINLKNLYLAWNCYFNKV  
CEKTNIEDGVFETLTNLELLSLSFNSLSHVPPKLPSSLRKLFLSNTQIKYISEEDFKGLINL  
TLDDLSGNCPRCFNAPFPVPCDGGASINIDRFAPQNLTLQLRYNLSSSTSLRKINAAWFKNM  
PHLKVLDELFNYLVGEIVSGAFLTMLPRLEILDLSFNYYIKGSYPQHINISRNFSKLLSLRAL  
HLRGYVFQELREDDFQPLMQPLNLTINLGINFIKQIDFKLFQNFNSNLEIIYLSNRI SPLV  
KDTQSYANSSSFQRHIRKRRSTDFEFDPHSNFYHFTRPLIKPQCAAYGKALDLSLNSIFFI  
GPNQFENLPDIACLNLSANSNAQVLSGTEFSAIPHVKYLDLTNNRLDFDNASALTELS DLEV  
LDLSYNSHYFRIAGVTHHLEFIQNFTNLKVLNLSHNNIYTLTDKYNLESKSLVELVFSGNRL  
DILWDDDNRYISIFKGLKNLTRLDLSLNRKHIPNEAFLNLPASLTE LHINDNMLKFFNWT  
LLQQFPRELLDLRGNKLLFLTDSLSDFTSSLRTLLLSHNRI SHLPSGFLSEVSSLKHLDLS  
SNLLKTINKSALETKTTTTKLSMLELHGPNPFECTCDIGDFRRWMDEHLNVKIPRLVDVICASP  
GDQRGKSIVSLELTTCVSDVTAVILFFFTFFITTMVLAALAHHLFYWDVWFIYNVCLAKVK  
GYRSLSTSQTfyDAYISYDTKDASVTDWVINElRYHLEESRDKNVLLCLEERDWDPLAIID  
NLMQSIHQSKTTFVVLTKKYAKSWNFKTAfYlALQRLMDENMDVlIFILLEPVLQHSQYLRL  
RQRICKSSILQWPDNPKAEGLFWQTLRNVVLTEndSRYNMYVDSIKQY

**Signal sequence:**

amino acids 1-26

**Transmembrane domain:**

amino acids 826-848

## FIGURE 212

CCAGGTCCAACCTGCACCTCGGTTCTATCGATTGAATTCCTCCCGGGGATCCTCTAGAGATCCCT  
CGACCTCGACCCACGCGTCCGCCAAGCTGGCCCTGCACGGCTGCAAGGAGGCTCCTGTGGA  
CAGGCCAGGCAGGTGGGCCTCAGGAGGTGCCTCCAGGCGGCCAGTGGGCCTGAGGCCCCAGC  
AAGGGCTAGGGTCCATCTCCAGTCCAGGACACAGCAGCGGCCACCATGGCCACGCTTGGGC  
TCCAGCAGCATCAGCAGCCCCAGGACCGGGGAGGCACAGGTGGCCCCACCAACCCGGAGGA  
GCAGTCTCTGCCCCGTGTCGCGGGGATGACTGATTCTCCTCCGCCAGGCCACCCAGAGGAGAA  
GGCCACCCCGCCTGGAGGACACAGGCCATGAGGGGCTCTCAGGAGGTGCTGCTGATGTGGCTT  
CTGGTGTTGGCAGTGGGCGGCACAGAGCACGCCTACCGGCCCGGCCGTAGGGTGTGTGCTGT  
CCGGGCTCACGGGACCCCTGTCTCCGAGTCGTTTCGTGCAGCGTGTGTACCAGCCCTTCCTCA  
CCACCTGCGACGGGCACCGGGCCTGCAGCACCTACCGAACCATCTATAGGACCGCCTACCGC  
CGCAGCCCTGGGCTGGCCCTGCCAGGCCTCGCTACGCGTGTGCCCCGGCTGGAAGAGGAC  
CAGCGGGCTTCTTGGGGCCTGTGAGCAGCAATATGCCAGCCGCATGCCGAACGGAGGGA  
GCTGTGTCCAGCCTGGCCGCTGCCGCTGCCCTGCAGGATGGCGGGGTGACACTTGCCAGTCA  
GATGTGGATGAATGCAGTGC TAGGAGGGGCGGCTGTCCCAGCGCTGCATCAACACCGCCCG  
CAGTTACTGGTGCCAGTGTGTTGGGAGGGGCACAGCCTGTCTGCAGACGTTACACTCTGTGTGC  
CCAAGGGAGGGCCCCCAGGGTGGCCCCAACCCGACAGGAGTGGACAGTGCATGAAGAGAA  
GAAGTGCAGAGGCTGCAGTCCAGGGTGGACCTGTGAGGAGAGAAGCTGCAGCTGGTGTGGC  
CCCACTGCACAGCCTGGCCTCGCAGGCACTGGAGCATGGGCTCCCGGACCCCGGCAGCCTCC  
TGGTGCACCTCTTCCAGCAGCTCGGCCGCATCGACTCCCTGAGCGAGCAGATTCTCTTCTTG  
GAGGAGCAGCTGGGGTCTGCTCCTGCAAGAAAGACTCGTGACTGCCAGCGCCCCAGGCTG  
GACTGAGCCCCTCACGCCGCCCTGCAGCCCCCATGCCCTGCCCAACATGTGGGGGTCCAG  
AAGCCACCTCGGGGTGACTGAGCGGAAGGCCAGGCAGGGCCTTCTCTCTTTCCTCTCCCC  
TTCTTCGGGAGGCTCCCCAGACCTGGCATGGGATGGGCTGGGATCTTCTCTGTGAATCCAC  
CCCTGGCTACCCCCACCTTGGCTACCCCAACGGCATCCCAAGGCCAGGTGGGCCCTCAGCTG  
AGGGAAGTACAGAGTCTCCCTGCTGGAGCCTGGGACCATGGCACAGGCCAGGCAGCCCCGAG  
GCTGGGTGGGGCCTCAGTGGGGCTGCTGCCTGACCCCGCAGCACATAAAATGAACGTGA  
AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGGGCGGCCCGCAGCTCTAGAGT  
CGACCTGCAGAAGCTTGGCCGCCATGGCCAACTTGTTATTGCAGCTTATAATGGTTACAAAT

## FIGURE 213

MRGSQEVLLMWLLVLAVGGTEHAYRPGRRVCAVRAHGDVSESFVQRVYQPFLTTCDGHRAC  
STYRTIYRTAYRRSPGLAPARPRYACCPGWKRTSGLPGACGAAICPPPCRNGGSCVQPGRCR  
CPAGWRGDTQCSDVDECSARRGGCPQRCINTAGSYWCQCWEGHSLSADGTLCVPKGGPPRVA  
PNPTGVDSAMKEEVQRLQSRVDLLEEKQLVLAPLHSLASQALEHGLPDPGSLLVHSFQQLG  
RIDSLSEQISFLEEQLGSCSCKKDS

Signal sequence:

1-19

1007031-102401

## **FIGURE 214**

GCCAGGCAGGTGGGCCTCAGGAGGTGCCTCCAGGCGGCCAGTGGGCCTGAGGCCCCAGCAAG  
GGCTAGGGTCCATCTCCAGTCCCAGGACACAGCAGCGGCCACCATGGCCACGCCTGGGCTCC  
AGCAGCATCAGAGCAGCCCCTGTGGTTGGCAGCAAAGTTAGCTTGGCTGGGCCCGCTGTGA  
GGGGCTTCGCGCTACGCCCTGCGGTGTCCCGAGGGGTGAGGTCTCCTCATCTTCTCCCTAGC  
AGTGGATGAGCAACCCAAACGGGGGCCCGGGAGGGGAACTGGCCCCGAGGGAGAGGAACCC  
AAAGCCACATCTGTAGCCAGGATGAGCAGTGTGAATCCAGGCAGCCCCAGGACCGGGGAGG  
CACAGGTGGCCCCCACCACCCGGAGGAGCAGCTCTGCCCCCTGTCCGGGGGATGACTGATTC  
TCCTCCGCCAGGCCACCCAGAGGAGAAGGCCACCCCGCCTGGAGGCACAGGCCATGAGGGGC  
TCTCAGGAGGTGCTGCTGATGTGGCTTCTGGTGTGGCAGTGGGCGGCACAGAGCACGCCTA  
CCGGCCCGGCCGTAGGGTGTGTGCTGTCCGGGCTCACGGGGACCCTGTCTCCGAGTCGTTG  
TGCAGCGTGTGTACACGCCCTTCTCACACCTGCGACGGGCACCGGGCCTGCAGCACCTAC  
CGAACCATCTATAGGACCGCCTACCGCCGAGCCCTGGGCTGGCCCCTGCCAGGCCTCGTCA  
CGCGTGCTGCCCCGGCTGGAAGAGGACCAGCGGGCTTCTGGGGCCTGTGGAGCAGCAATAT  
GCCAGCCGCATGCCGGAACGGAGGGAGCTGTGTCCAGCCTGGCCGCTGCCGCTGCCCTGCA  
GGATGGCGGGTGACACTTGCCAGTCAGATGTGGATGAATGCAGTGTAGGAGGGGCGGCTG  
TCCCCAGCGCTGCATCAACCCGCCGAGTTACTGGTGCCAGTGTGGGAGGGGCACAGCC  
TGTCTGCAGACGCTACACTCTGTGTGCCAAGGGAGGGCCCCCAGGGTGGCCCCCAACCCG  
ACAGGAGTGGACAGTGAATGAAGGAAGAAGTGACAGAGCTGCAGTCCAGGGTGGACCTGCT  
GGAGGAGAAGCTGCAGCTGGTGCTGGCCCCACTGCACAGCCTGGCCTCGCAGGCACTGGAGC  
ATGGGCTCCCCGACCCCGGAGCCTCCTGGTGCACTCCTTCCAGCAGCTCGGCCGATCGAC  
TCCCTGAGCGAGCAGATTTCTTCTTGAGGAGCAGCTGGGGTCTGCTCTGCAAGAAAGA  
CTCGTGACTGCCCAGCGCTCCAGGCTGGACTGAGCCCCCTACGCCGCCCTGCAGCCCCCATG  
CCCTGCCCAACATGCTGGGGTCCAGAAGCCACCTCGGGGTGACTGAGCGGAAGGCCAGGC  
AGGGCCTTCTCTCTTCTCTCTCTCTCGGGAGGCTCCCAGACCTGGCATGGGAT  
GGGCTGGGATCTTCTGTGAATCCACCCCTGGCTACCCCCACCCTGGCTACCCCAACGGCA  
TCCCAAGGCAGGTGGACCTCAGCTGAGGGAAGGTACGAGCTCCCTGCTGGAGCCTGGGAC  
CCATGGGCACAGGCCAGGCAGCCCGAGGCTGGGTGGGGCCTCAGTGGGGGTGCTGCCTGAC  
CCCCAGCACAAATAAAATGAAACGTG

## **FIGURE 215**

MRGSQEVLLMWLLVLAVGGTEHAYRPGRRVCAVRAHGDVSESFVQRVYQPFLTTCDGHRAC  
STYRTIYRTAYRRSPGLAPARPRYACCPGWKRTSGLPGACGAAICQPPCRNGGSCVQPGRCR  
CPAGWRGDTQCSDVDECSARRGGCPQRCINTAGSYWCQCWEGHSLSADGTL CVPKGGPPRVA  
PNPTGVDSAMKEEVQRLQSRVDLLEEKQLVLAPLHSLASQALEHGLPDPGSLLVHSFQQLG  
RIDSLSSEQISFLEEQLGSCSCKKDS

**Signal sequence:**

1-19

10017881-102401

## FIGURE 216

CCCACGCGTCCGAAGCTGGCCCTGCACGGCTGCAAGGGAGGCTCCTGTGGACAGGCCAGGCA  
GGTGGGCCTCAGGAGGTGCCTCCAGGCGGCCAGTGGGCCTGAGGCCCCAGCAAGGGCTAGGG  
TCCATCTCCAGTCCAGGACACAGCAGCGGCCACCATGGCCACGCTTGGGCTCCAGCAGCAT  
CAGCAGCCCCCAGGACCGGGGAGGCAAGGTGGCCCCCACCACCCGGAGGAGCAGCTCCTGC  
CCCTGTCCGGGGATGACTGATTCTCCTCCGCCAGGCCACCAGAGGAGAAGGCCACCCCGC  
CTGGAGGCACAGGCCATGAGGGGCTCTCAGGAGGTGCTGCTGATGTGGCTTCTGGTGTGGC  
AGTGGGCGGCACAGAGCACGCCTACCGGCCCGGCCGTAGGGTGTGTGCTGTCCGGGCTCACG  
GGGACCCTGTCTCCGAGTCGTTCTGTGCAGCGTGTGTACCAGCCCTTCCTCACCACTTGCAGC  
GGGCACCGGCCTGCAGCACTTACCGAACCATCTATAGGACCGCCTACCGCCGAGCCCTGG  
GCTGGCCCCCTGCCAGGCCTCGCTACGCGTGTGCTGCCCGGCTGGAAGAGGACCAGCGGGCTTC  
CTGGGGCCTGTGGAGCAGCAATATGCCAGCCGCCATGCCGGAACGAGGGAGCTGTGTCCAG  
CCTGGCCGCTGCCGCTGCCCTGCAGGATGGCGGGGTGACACTTGCCAGTCAGATGTGGATGA  
ATGCAGTGCTAGGAGGGGCGGCTGTCCCCAGCGCTGCGTCAACACCGCGGCGAGTTACTGGT  
GCCAGTGTGGGAGGGGCACAGCCTGTCTGCAGACGGTACACTCTGTGTGCCCAAGGGAGGG  
CCCCCAGGGTGGCCCCCAACCGACAGGAGTGGACAGTGCAATGAAGGAAGAAGTGACAGAG  
GCTGCAGTCCAGGGTGGACCTGCTGGAGGAGAAGCTGCAGCTGGTGTGGCCCCACTGCACA  
GCCTGGCCTCGCAGGCACTGGAGCATGGGCTCCCGGACCCCGGAGCCTCCTGGTGCCTCC  
TTCCAGCAGCTCGGCCGCATCGACTCCCTGAGCGAGCAGATTTCCTTCTGGAGGAGCAGCT  
GGGGTCTGTCTCTGCAAGAAAGACTCGTGACTGCCCCAGCGCCCCAGGCTGGACTGAGCCCC  
TCACGCCGCCCTGCAGCCCCATGCCCTGCCAACATGCTGGGGTCCAGAAGCCACCTCG  
GGGTGACTGAGCGGAAGGCCAGGCAGGGCCTTCTCTCTTCTCTCTCCCTTCTCTGGGAG  
GCTCCCCAGACCTTGGCATGGGATGGGCTGGGATCTTCTCTGTGAATCCACCCCTGGCTACC  
CCCACCTTGCTACCCCAACGGCATCCCAAGGCCAGGTGGGCCCTCAGCTGAGGGAAGGTAC  
GAGCTCCCTGTCTGGAGCCTGGGACCCATGGCACAGGCCAGGCAGCCCGGAGGCTGGGTGGGG  
CCTCAGTGGGGGCTGTGCCTGACCCCCAGCACATAAAAAATGAAACGTG

## FIGURE 217

MRGSQEVLLMWLLVLAVGGTEHAYRPGRRVCAVRAHGDPVSESFVQRVYQPFLTTCDGHRAC  
STYRTIYRTAYRRSPGLAPARPRYACCPGWKRTSGLPGACGAACQPPCRNGGSCVPGRRCR  
CFAGWRGDTCCQSDVDECSARRGGCFQRCVNTAGSYWCQCWEGHSLSADGTLCVPKGGPFRVA  
FNPTGVDSAMKEEVQRLQSRVDLLEEKQLVLAPLHSLASQALEHGLPDPGSLLVHSFQQLG  
RIDSLSSEQISFLEEQLGSCSCKKDS

**Signal sequence:**

1-19

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
841  
842  
843  
844  
845  
846  
847  
848  
849  
850  
851  
852  
853  
854  
855  
856  
857  
858  
859  
860  
861  
862  
863  
864  
865  
866  
867  
868  
869  
870  
871  
872  
873  
874  
875  
876  
877  
878  
879  
880  
881  
882  
883  
884  
885  
886  
887  
888  
889  
890  
891  
892  
893  
894  
895  
896  
897  
898  
899  
900  
901  
902  
903  
904  
905  
906  
907  
908  
909  
910  
911  
912  
913  
914  
915  
916  
917  
918  
919  
920  
921  
922  
923  
924  
925  
926  
927  
928  
929  
930  
931  
932  
933  
934  
935  
936  
937  
938  
939  
940  
941  
942  
943  
944  
945  
946  
947  
948  
949  
950  
951  
952  
953  
954  
955  
956  
957  
958  
959  
960  
961  
962  
963  
964  
965  
966  
967  
968  
969  
970  
971  
972  
973  
974  
975  
976  
977  
978  
979  
980  
981  
982  
983  
984  
985  
986  
987  
988  
989  
990  
991  
992  
993  
994  
995  
996  
997  
998  
999  
1000  
1001  
1002  
1003  
1004  
1005  
1006  
1007  
1008  
1009  
1010  
1011  
1012  
1013  
1014  
1015  
1016  
1017  
1018  
1019  
1020  
1021  
1022  
1023  
1024  
1025  
1026  
1027  
1028  
1029  
1030  
1031  
1032  
1033  
1034  
1035  
1036  
1037  
1038  
1039  
1040  
1041  
1042  
1043  
1044  
1045  
1046  
1047  
1048  
1049  
1050  
1051  
1052  
1053  
1054  
1055  
1056  
1057  
1058  
1059  
1060  
1061  
1062  
1063  
1064  
1065  
1066  
1067  
1068  
1069  
1070  
1071  
1072  
1073  
1074  
1075  
1076  
1077  
1078  
1079  
1080  
1081  
1082  
1083  
1084  
1085  
1086  
1087  
1088  
1089  
1090  
1091  
1092  
1093  
1094  
1095  
1096  
1097  
1098  
1099  
1100  
1101  
1102  
1103  
1104  
1105  
1106  
1107  
1108  
1109  
1110  
1111  
1112  
1113  
1114  
1115  
1116  
1117  
1118  
1119  
1120  
1121  
1122  
1123  
1124  
1125  
1126  
1127  
1128  
1129  
1130  
1131  
1132  
1133  
1134  
1135  
1136  
1137  
1138  
1139  
1140  
1141  
1142  
1143  
1144  
1145  
1146  
1147  
1148  
1149  
1150  
1151  
1152  
1153  
1154  
1155  
1156  
1157  
1158  
1159  
1160  
1161  
1162  
1163  
1164  
1165  
1166  
1167  
1168  
1169  
1170  
1171  
1172  
1173  
1174  
1175  
1176  
1177  
1178  
1179  
1180  
1181  
1182  
1183  
1184  
1185  
1186  
1187  
1188  
1189  
1190  
1191  
1192  
1193  
1194  
1195  
1196  
1197  
1198  
1199  
1200  
1201  
1202  
1203  
1204  
1205  
1206  
1207  
1208  
1209  
1210  
1211  
1212  
1213  
1214  
1215  
1216  
1217  
1218  
1219  
1220  
1221  
1222  
1223  
1224  
1225  
1226  
1227  
1228  
1229  
1230  
1231  
1232  
1233  
1234  
1235  
1236  
1237  
1238  
1239  
1240  
1241  
1242  
1243  
1244  
1245  
1246  
1247  
1248  
1249  
1250  
1251  
1252  
1253  
1254  
1255  
1256  
1257  
1258  
1259  
1260  
1261  
1262  
1263  
1264  
1265  
1266  
1267  
1268  
1269  
1270  
1271  
1272  
1273  
1274  
1275  
1276  
1277  
1278  
1279  
1280  
1281  
1282  
1283  
1284  
1285  
1286  
1287  
1288  
1289  
1290  
1291  
1292  
1293  
1294  
1295  
1296  
1297  
1298  
1299  
1300  
1301  
1302  
1303  
1304  
1305  
1306  
1307  
1308  
1309  
1310  
1311  
1312  
1313  
1314  
1315  
1316  
1317  
1318  
1319  
1320  
1321  
1322  
1323  
1324  
1325  
1326  
1327  
1328  
1329  
1330  
1331  
1332  
1333  
1334  
1335  
1336  
1337  
1338  
1339  
1340  
1341  
1342  
1343  
1344  
1345  
1346  
1347  
1348  
1349  
1350  
1351  
1352  
1353  
1354  
1355  
1356  
1357  
1358  
1359  
1360  
1361  
1362  
1363  
1364  
1365  
1366  
1367  
1368  
1369  
1370  
1371  
1372  
1373  
1374  
1375  
1376  
1377  
1378  
1379  
1380  
1381  
1382  
1383  
1384  
1385  
1386  
1387  
1388  
1389  
1390  
1391  
1392  
1393  
1394  
1395  
1396  
1397  
1398  
1399  
1400  
1401  
1402  
1403  
1404  
1405  
1406  
1407  
1408  
1409  
1410  
1411  
1412  
1413  
1414  
1415  
1416  
1417  
1418  
1419  
1420  
1421  
1422  
1423  
1424  
1425  
1426  
1427  
1428  
1429  
1430  
1431  
1432  
1433  
1434  
1435  
1436  
1437  
1438  
1439  
1440  
1441  
1442  
1443  
1444  
1445  
1446  
1447  
1448  
1449  
1450  
1451  
1452  
1453  
1454  
1455  
1456  
1457  
1458  
1459  
1460  
1461  
1462  
1463  
1464  
1465  
1466  
1467  
1468  
1469  
1470  
1471  
1472  
1473  
1474  
1475  
1476  
1477  
1478  
1479  
1480  
1481  
1482  
1483  
1484  
1485  
1486  
1487  
1488  
1489  
1490  
1491  
1492  
1493  
1494  
1495  
1496  
1497  
1498  
1499  
1500  
1501  
1502  
1503  
1504  
1505  
1506  
1507  
1508  
1509  
1510  
1511  
1512  
1513  
1514  
1515  
1516  
1517  
1518  
1519  
1520  
1521  
1522  
1523  
1524  
1525  
1526  
1527  
1528  
1529  
1530  
1531  
1532  
1533  
1534  
1535  
1536  
1537  
1538  
1539  
1540  
1541  
1542  
1543  
1544  
1545  
1546  
1547  
1548  
1549  
1550  
1551  
1552  
1553  
1554  
1555  
1556  
1557  
1558  
1559  
1560  
1561  
1562  
1563  
1564  
1565  
1566  
1567  
1568  
1569  
1570  
1571  
1572  
1573  
1574  
1575  
1576  
1577  
1578  
1579  
1580  
1581  
1582  
1583  
1584  
1585  
1586  
1587  
1588  
1589  
1590  
1591  
1592  
1593  
1594  
1595  
1596  
1597  
1598  
1599  
1600  
1601  
1602  
1603  
1604  
1605  
1606  
1607  
1608  
1609  
1610  
1611  
1612  
1613  
1614  
1615  
1616  
1617  
1618  
1619  
1620  
1621  
1622  
1623  
1624  
1625  
1626  
1627  
1628  
1629  
1630  
1631  
1632  
1633  
1634  
1635  
1636  
1637  
1638  
1639  
1640  
1641  
1642  
1643  
1644  
1645  
1646  
1647  
1648  
1649  
1650  
1651  
1652  
1653  
1654  
1655  
1656  
1657  
1658  
1659  
1660  
1661  
1662  
1663  
1664  
1665  
1666  
1667  
1668  
1669  
1670  
1671  
1672  
1673  
1674  
1675  
1676  
1677  
1678  
1679  
1680  
1681  
1682  
1683  
1684  
1685  
1686  
1687  
1688  
1689  
1690  
1691  
1692  
1693  
1694  
1695  
1696  
1697  
1698  
1699  
1700  
1701  
1702  
1703  
1704  
1705  
1706  
1707  
1708  
1709  
1710  
1711  
1712  
1713  
1714  
1715  
1716  
1717  
1718  
1719  
1720  
1721  
1722  
1723  
1724  
1725  
1726  
1727  
1728  
1729  
1730  
1731  
1732  
1733  
1734  
1735  
1736  
1737  
1738  
1739  
1740  
1741  
1742  
1743  
1744  
1745  
1746  
1747  
1748  
1749  
1750  
1751  
1752  
1753  
1754  
1755  
1756  
1757  
1758  
1759  
1760  
1761  
1762  
1763  
1764  
1765  
1766  
1767  
1768  
1769  
1770  
1771  
1772  
1773  
1774  
1775  
1776  
1777  
1778  
1779  
1780  
1781  
1782  
1783  
1784  
1785  
1786  
1787  
1788  
1789  
1790  
1791  
1792  
1793  
1794  
1795  
1796  
1797  
1798  
1799  
1800  
1801  
1802  
1803  
1804  
1805  
1806  
1807  
1808  
1809  
1810  
1811  
1812  
1813  
1814  
1815  
1816  
1817  
1818  
1819  
1820  
1821  
1822  
1823  
1824  
1825  
1826  
1827  
1828  
1829  
1830  
1831  
1832  
1833  
1834  
1835  
1836  
1837  
1838  
1839  
1840  
1841  
1842  
1843  
1844  
1845  
1846  
1847  
1848  
1849  
1850  
1851  
1852  
1853  
1854  
1855  
1856  
1857  
1858  
1859  
1860  
1861  
1862  
1863  
1864  
1865  
1866  
1867  
1868  
1869  
1870  
1871  
1872  
1873  
1874  
1875  
1876  
1877  
1878  
1879  
1880  
1881  
1882  
1883  
1884  
1885  
1886  
1887  
1888  
1889  
1890  
1891  
1892  
1893  
1894  
1895  
1896  
1897  
1898  
1899  
1900  
1901  
1902  
1903  
1904  
1905  
1906  
1907  
1908  
1909  
1910  
1911  
1912  
1913  
1914  
1915  
1916  
1917  
1918  
1919  
1920  
1921  
1922  
1923  
1924  
1925  
1926  
1927  
1928  
1929  
1930  
1931  
1932  
1933  
1934  
1935  
1936  
1937  
1938  
1939  
1940  
1941  
1942  
1943  
1944  
1945  
1946  
1947  
1948  
1949  
1950  
1951  
1952  
1953  
1954  
1955  
1956  
1957  
1958  
1959  
1960  
1961  
1962  
1963  
1964  
1965  
1966  
1967  
1968  
1969  
1970  
1971  
1972  
1973  
1974  
1975  
1976  
1977  
1978  
1979  
1980  
1981  
1982  
1983  
1984  
1985  
1986  
1987  
1988  
1989  
1990  
1991  
1992  
1993  
1994  
1995  
1996  
1997  
1998  
1999  
2000  
2001  
2002  
2003  
2004  
2005  
2006  
2007  
2008  
2009  
2010  
2011  
2012  
2013  
2014  
2015  
2016  
2017  
2018  
2019  
2020  
2021  
2022  
2023  
2024  
2025  
2026  
2027  
2028  
2029  
2030  
2031  
2032  
2033  
2034  
2035  
2036  
2037  
2038  
2039  
2040  
2041  
2042  
2043  
2044  
2045  
2046  
2047  
2048  
2049  
2050  
2051  
2052  
2053  
2054  
2055  
2056  
2057  
2058  
2059  
2060  
2061  
2062  
2063  
2064  
2065  
2066  
2067  
2068  
2069  
2070  
2071  
2072  
2073  
2074  
2075  
2076  
2077  
2078  
2079  
2080  
2081  
2082  
2083  
2084  
2085  
2086  
2087  
2088  
2089  
2090  
2091  
2092  
2093  
2094  
2095  
2096  
2097  
2098  
2099  
2100  
2101  
2102  
2103  
2104  
2105  
2106  
2107  
2108  
2109  
2110  
2111  
2112  
2113  
2114  
2115  
2116  
2117  
2118  
2119  
2120  
2121  
2122  
2123  
2124  
2125  
2126  
2127  
2128  
2129  
2130  
2131  
2132  
2133  
2134  
2135  
2136  
2137  
2138  
2139  
2140  
2141  
2142  
2143  
2144  
2145  
2146  
2147  
2148  
2149  
2150  
2151  
2152  
2153  
2154  
2155  
2156  
2157  
2158  
2159  
2160  
2161  
2162  
2163  
2164  
2165  
2166  
2167  
2168  
2169  
2170  
2171  
2172  
2173  
2174  
2175  
2176  
2177  
2178  
2179  
2180  
2181  
2182  
2183  
2184  
2185  
2186  
2187  
2188  
2189  
2190  
2191  
2192

Year	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035	2036	2037	2038	2039	2040	2041	2042	2043	2044	2045	2046	2047	2048	2049	2050	2051	2052	2053	2054	2055	2056	2057	2058	2059	2060	2061	2062	2063	2064	2065	2066	2067	2068	2069	2070	2071	2072	2073	2074	2075	2076	2077	2078	2079	2080	2081	2082	2083	2084	2085	2086	2087	2088	2089	2090	2091	2092	2093	2094	2095	2096	2097	2098	2099	2100
1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035	2036	2037	2038	2039	2040	2041	2042	2043	2044	2045	2046	2047	2048	2049	2050	2051	2052	2053	2054	2055	2056	2057	2058	2059	2060	2061	2062	2063	2064	2065	2066	2067	2068	2069	2070	2071	2072	2073	2074	2075	2076	2077	2078	2079	2080	2081	2082	2083	2084	2085	2086	2087	2088	2089	2090	2091	2092	2093	2094	2095	2096	2097	2098	2099	2100	

[illegible]

## **FIGURE 219**

MSVMVVRKKVTRKWEKLPGRNTFCDDGRVMMARQKGFYLTFLILGTCTLFFAFECRYLAV  
QLSPAIPVFAMLFLLFSMATLLRTSFSDPGVIPRALPDEAAFIEMEIEATNGAVPQGQRPPP  
RIKNFQINNQIVKLKYCYTCKIFRPPRASHCSICDNCVERFDHHCPCWVGNCVGNRYRYFYL  
FILSLSLLTIIYVFAFNIVYVALKSLKIGFLETCLKETPGTVLEVLCFFTLWSVVGLTGFHTF  
LVALNQTTNEDIKGSWTGKNRVQNPYSHGNIVKNCCEVLCGPLPPSVLDRRGILPLEESGSR  
PPSTQETSSSLLPQSPAPTEHLNSNEMPEDSSTPEEMPPPEPPEPPQEAAAEK

**Putative transmembrane domains:**

amino acids 36-55 (type II TM), 65-84, 188-208, 229-245

bioRxiv preprint doi: <https://doi.org/10.1101/000000>; this version posted January 1, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

## **FIGURE 220**

AAAACCTGTATTTTTTACAATGCAAATAGACAATNANCCTGGAGGTCTTTGAATTAGGTAT  
TATAGGGATGGTGGGGTTGATTTTTNTTCCTGGAGGCTTTTGGCTTTGGACTCTCNCTTTCT  
CCACAGAGCNCTTCGACCATCACTGCCCCCTGGGTGGGAATTGTGTTGGAAGAGGAACTA  
CCGCTANTTCTACCTCTTCATCCTTTNTCTCTCCNCCTCACAATCTATGTCTTCGCCTTCA  
ACATCGT

2020-2021-2022

## **FIGURE 221**

GTGTGTCTTCAGCAAAACAGTGGATTAAATCTCCTTGCAACAAGCTTGAGAGCAACACAA  
TCTATCAGGAAAGAAAGAAAAAAACCGAACCTGACAAAAAGAGAAAAAGAGA  
AAAAAATCATGAAAACCATCCAGCCAAAAATGCACAATCTATCTCTTGGGCAATCTTCAC  
GGGGCTGGCTGCTCTGTGTCTCTTCCAAGGAGTGCCCGTGCGCAGCGGAGATGCCACCTTCC  
CCAAAGCTATGGACAACGTGACGGTCCGGCAGGGGAGAGCGCCACCTCAGGTGCACTATT  
GACAAACCGGGTCACCCGGGTGGCTGGCTAAACCGCAGCACCATCCTCTATGCTGGGAATGA  
CAAGTGGTGCCTGGATCCTCGCGTGGTCTTCTGAGCAACACCCAAACGCAGTACAGCATCG  
AGATCCAGAACGTGGATGTGTATGACGAGGGCCCTTACACCTGCTCGGTGCAGACAGACAA  
CACCCAAAGACCTCTAGGGTCCACCTCATTGTGCAAGTATCTCCCAAAATTGTAGAGATTT  
TTCAGATATCTCCATTAATGAAGGGAACAATATTAGCCTCACCTGCATAGCAACTGGTAGAC  
CAGAGCCTACGGTTACTTGGAGACACATCTCTCCAAAGCGGTTGGCTTTGTGAGTGAAGAC  
GAATACTTGGAAATTCAGGGCATCACCCGGGAGCAGTCAGGGGACTACGAGTGCAGTGCCTC  
CAATGACGTGGCCGCGCCCGTGGTACGGAGAGTAAGGTCACCGTGAACCTATCCACCATACA  
TTTCAGAAGCCAAAGGTACAGGTGTCCCCGTGGGACAAAAGGGGACACTGCACTGTGAAGCC  
TCAGCAGTCCCCTCAGCAGAATTCAGTGGTACAAGGATGACAAAAGACTGATTGAAGGAAA  
GAAAGGGGTGAAAGTGAAAACAGACCTTTCTCTCAAACTCATCTTCTTCAATGTCTCTG  
AACATGACTATGGAACTACACTTGCCTGGCCTCCAACAAGCTGGGCCACACCAATGCCAGC  
ATCATGCTATTTGGTCCAGGCGCCGTGAGCGAGGTGAGCAACGGCAGCTCGAGGAGGGCAGG  
CTGCGTCTGGCTGCTGCCTCTTCTGGTCTTGCACTGCTTCTCAAATTTTGATGTGAGTGCC  
ACTTCCCCACCCGGGAAAGGTGCGGCCACCACCACCACCAACACAACAGCAATGGCAACAC  
CGACAGCAACCAATCAGATATATACAAATGAAATTAGAAGAAACACAGCCTCATGGGACAGA  
AATTTGAGGGAGGGGAACAAAGAATACTTTGGGGGAAAGAGTTTTAAAAAGAAATTGAA  
AATTGCCTTGCAGATATTTAGGTACAATGGAGTTTCTTTTCCAAACGGGAAGAACACAGC  
ACACCCGGCTTGAGCCCACTGCAAGCTGCATCGTGCAACCTCTTTGGTGCCAGTGTGGGCAA  
GGGCTCAGCCTCTCTGCCACAGAGTGCCCCACGTGGAACATTCTGGAGCTGGCCATCCCA  
AATTCATCAGTCCATAGAGACGAACAGAAATGAGACCTTCCGGCCCAAGCGTGGCGCTGCGG  
GCACCTTGGTAGACTGTGCCACCACGGCGTGTGTTGTGAACGTGAAATAAAAAGAGCAAAA  
AAAAA

## **FIGURE 222**

MKTIQPKMHNSISWAIFTGLAALCLFQGVFVRSGDATFPKAMDNVTVRQGESATLRCTIDNR  
VTRVAWLNRSITILYAGNDKWCLDPRVVLLSNTQTQYSIEIQNVDVYDEGPYTCSVQTDNHPK  
TSRVHLIVQVSPKIVEISSDISINEGNNISLTCIATGRPEPTVTWRHISPKAVGFVSEDEYL  
EIQGITREQSGDYECSASNDVAAPVVRVKVTVNYPPISEAKGTGVPVGQKGTLCQCEASAV  
PSAEFQWYKDDKRLIEGKKGVKVENRPFLSKLIFFNVSEHDYGNITCVASNKLGHNTNASIML  
FGPGAVSEVSNGTSRRAGCVWLLPLLVLHLLKF

**Signal peptide:**

amino acids 1-28

1001081 1024001

## FIGURE 223

GAAAAAAATCATGAAAACCATCCAGCCAAAAATGCACAATTCTATCTCTTGGGCAATCTTC  
ACGGGGCTGGCTGCTCTGTGTCTCTTCCAAGGAGTGCCCGTGCGCAGCGGAGATGCCACCTT  
CCCCAAAGCTATGGACAACGTGACGGTCCGGCAGGGGGAGAGCGCCACCCTCAGGTGCACTA  
TTGACAACCGGGTCACCCGGGTGGCCTGGCTAAACCGCAGCACCATCCTCTATGCTGGGAAT  
GACAAGTGGTGCCCTGGATCCTCGCGTGGTCCTTCTGAGCAACACCCAAACGCAGTACAGCAT  
CGAGATCCAGAACGTGGATGTGTATGACGAGGGCCCTTACACCTGCTCGGTGCAGACAGACA  
ACCACCCAAAGACCTCTAGGGTCCACCTCATTGTGCAAGTATCTCCCAAAATTGTAGAGATT  
TCTTCAGATATCTCCATTAAATGAAGGGAACAATATTAGCCTCACCTGCATAGCAACTGGTAG  
ACCAGAG

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
841  
842  
843  
844  
845  
846  
847  
848  
849  
850  
851  
852  
853  
854  
855  
856  
857  
858  
859  
860  
861  
862  
863  
864  
865  
866  
867  
868  
869  
870  
871  
872  
873  
874  
875  
876  
877  
878  
879  
880  
881  
882  
883  
884  
885  
886  
887  
888  
889  
890  
891  
892  
893  
894  
895  
896  
897  
898  
899  
900  
901  
902  
903  
904  
905  
906  
907  
908  
909  
910  
911  
912  
913  
914  
915  
916  
917  
918  
919  
920  
921  
922  
923  
924  
925  
926  
927  
928  
929  
930  
931  
932  
933  
934  
935  
936  
937  
938  
939  
940  
941  
942  
943  
944  
945  
946  
947  
948  
949  
950  
951  
952  
953  
954  
955  
956  
957  
958  
959  
960  
961  
962  
963  
964  
965  
966  
967  
968  
969  
970  
971  
972  
973  
974  
975  
976  
977  
978  
979  
980  
981  
982  
983  
984  
985  
986  
987  
988  
989  
990  
991  
992  
993  
994  
995  
996  
997  
998  
999  
1000  
1001  
1002  
1003  
1004  
1005  
1006  
1007  
1008  
1009  
1010  
1011  
1012  
1013  
1014  
1015  
1016  
1017  
1018  
1019  
1020  
1021  
1022  
1023  
1024  
1025  
1026  
1027  
1028  
1029  
1030  
1031  
1032  
1033  
1034  
1035  
1036  
1037  
1038  
1039  
1040  
1041  
1042  
1043  
1044  
1045  
1046  
1047  
1048  
1049  
1050  
1051  
1052  
1053  
1054  
1055  
1056  
1057  
1058  
1059  
1060  
1061  
1062  
1063  
1064  
1065  
1066  
1067  
1068  
1069  
1070  
1071  
1072  
1073  
1074  
1075  
1076  
1077  
1078  
1079  
1080  
1081  
1082  
1083  
1084  
1085  
1086  
1087  
1088  
1089  
1090  
1091  
1092  
1093  
1094  
1095  
1096  
1097  
1098  
1099  
1100  
1101  
1102  
1103  
1104  
1105  
1106  
1107  
1108  
1109  
1110  
1111  
1112  
1113  
1114  
1115  
1116  
1117  
1118  
1119  
1120  
1121  
1122  
1123  
1124  
1125  
1126  
1127  
1128  
1129  
1130  
1131  
1132  
1133  
1134  
1135  
1136  
1137  
1138  
1139  
1140  
1141  
1142  
1143  
1144  
1145  
1146  
1147  
1148  
1149  
1150  
1151  
1152  
1153  
1154  
1155  
1156  
1157  
1158  
1159  
1160  
1161  
1162  
1163  
1164  
1165  
1166  
1167  
1168  
1169  
1170  
1171  
1172  
1173  
1174  
1175  
1176  
1177  
1178  
1179  
1180  
1181  
1182  
1183  
1184  
1185  
1186  
1187  
1188  
1189  
1190  
1191  
1192  
1193  
1194  
1195  
1196  
1197  
1198  
1199  
1200  
1201  
1202  
1203  
1204  
1205  
1206  
1207  
1208  
1209  
1210  
1211  
1212  
1213  
1214  
1215  
1216  
1217  
1218  
1219  
1220  
1221  
1222  
1223  
1224  
1225  
1226  
1227  
1228  
1229  
1230  
1231  
1232  
1233  
1234  
1235  
1236  
1237  
1238  
1239  
1240  
1241  
1242  
1243  
1244  
1245  
1246  
1247  
1248  
1249  
1250  
1251  
1252  
1253  
1254  
1255  
1256  
1257  
1258  
1259  
1260  
1261  
1262  
1263  
1264  
1265  
1266  
1267  
1268  
1269  
1270  
1271  
1272  
1273  
1274  
1275  
1276  
1277  
1278  
1279  
1280  
1281  
1282  
1283  
1284  
1285  
1286  
1287  
1288  
1289  
1290  
1291  
1292  
1293  
1294  
1295  
1296  
1297  
1298  
1299  
1300  
1301  
1302  
1303  
1304  
1305  
1306  
1307  
1308  
1309  
1310  
1311  
1312  
1313  
1314  
1315  
1316  
1317  
1318  
1319  
1320  
1321  
1322  
1323  
1324  
1325  
1326  
1327  
1328  
1329  
1330  
1331  
1332  
1333  
1334  
1335  
1336  
1337  
1338  
1339  
1340  
1341  
1342  
1343  
1344  
1345  
1346  
1347  
1348  
1349  
1350  
1351  
1352  
1353  
1354  
1355  
1356  
1357  
1358  
1359  
1360  
1361  
1362  
1363  
1364  
1365  
1366  
1367  
1368  
1369  
1370  
1371  
1372  
1373  
1374  
1375  
1376  
1377  
1378  
1379  
1380  
1381  
1382  
1383  
1384  
1385  
1386  
1387  
1388  
1389  
1390  
1391  
1392  
1393  
1394  
1395  
1396  
1397  
1398  
1399  
1400  
1401  
1402  
1403  
1404  
1405  
1406  
1407  
1408  
1409  
1410  
1411  
1412  
1413  
1414  
1415  
1416  
1417  
1418  
1419  
1420  
1421  
1422  
1423  
1424  
1425  
1426  
1427  
1428  
1429  
1430  
1431  
1432  
1433  
1434  
1435  
1436  
1437  
1438  
1439  
1440  
1441  
1442  
1443  
1444  
1445  
1446  
1447  
1448  
1449  
1450  
1451  
1452  
1453  
1454  
1455  
1456  
1457  
1458  
1459  
1460  
1461  
1462  
1463  
1464  
1465  
1466  
1467  
1468  
1469  
1470  
1471  
1472  
1473  
1474  
1475  
1476  
1477  
1478  
1479  
1480  
1481  
1482  
1483  
1484  
1485  
1486  
1487  
1488  
1489  
1490  
1491  
1492  
1493  
1494  
1495  
1496  
1497  
1498  
1499  
1500  
1501  
1502  
1503  
1504  
1505  
1506  
1507  
1508  
1509  
1510  
1511  
1512  
1513  
1514  
1515  
1516  
1517  
1518  
1519  
1520  
1521  
1522  
1523  
1524  
1525  
1526  
1527  
1528  
1529  
1530  
1531  
1532  
1533  
1534  
1535  
1536  
1537  
1538  
1539  
1540  
1541  
1542  
1543  
1544  
1545  
1546  
1547  
1548  
1549  
1550  
1551  
1552  
1553  
1554  
1555  
1556  
1557  
1558  
1559  
1560  
1561  
1562  
1563  
1564  
1565  
1566  
1567  
1568  
1569  
1570  
1571  
1572  
1573  
1574  
1575  
1576  
1577  
1578  
1579  
1580  
1581  
1582  
1583  
1584  
1585  
1586  
1587  
1588  
1589  
1590  
1591  
1592  
1593  
1594  
1595  
1596  
1597  
1598  
1599  
1600  
1601  
1602  
1603  
1604  
1605  
1606  
1607  
1608  
1609  
1610  
1611  
1612  
1613  
1614  
1615  
1616  
1617  
1618  
1619  
1620  
1621  
1622  
1623  
1624  
1625  
1626  
1627  
1628  
1629  
1630  
1631  
1632  
1633  
1634  
1635  
1636  
1637  
1638  
1639  
1640  
1641  
1642  
1643  
1644  
1645  
1646  
1647  
1648  
1649  
1650  
1651  
1652  
1653  
1654  
1655  
1656  
1657  
1658  
1659  
1660  
1661  
1662  
1663  
1664  
1665  
1666  
1667  
1668  
1669  
1670  
1671  
1672  
1673  
1674  
1675  
1676  
1677  
1678  
1679  
1680  
1681  
1682  
1683  
1684  
1685  
1686  
1687  
1688  
1689  
1690  
1691  
1692  
1693  
1694  
1695  
1696  
1697  
1698  
1699  
1700  
1701  
1702  
1703  
1704  
1705  
1706  
1707  
1708  
1709  
1710  
1711  
1712  
1713  
1714  
1715  
1716  
1717  
1718  
1719  
1720  
1721  
1722  
1723  
1724  
1725  
1726  
1727  
1728  
1729  
1730  
1731  
1732  
1733  
1734  
1735  
1736  
1737  
1738  
1739  
1740  
1741  
1742  
1743  
1744  
1745  
1746  
1747  
1748  
1749  
1750  
1751  
1752  
1753  
1754  
1755  
1756  
1757  
1758  
1759  
1760  
1761  
1762  
1763  
1764  
1765  
1766  
1767  
1768  
1769  
1770  
1771  
1772  
1773  
1774  
1775  
1776  
1777  
1778  
1779  
1780  
1781  
1782  
1783  
1784  
1785  
1786  
1787  
1788  
1789  
1790  
1791  
1792  
1793  
1794  
1795  
1796  
1797  
1798  
1799  
1800  
1801  
1802  
1803  
1804  
1805  
1806  
1807  
1808  
1809  
1810  
1811  
1812  
1813  
1814  
1815  
1816  
1817  
1818  
1819  
1820  
1821  
1822  
1823  
1824  
1825  
1826  
1827  
1828  
1829  
1830  
1831  
1832  
1833  
1834  
1835  
1836  
1837  
1838  
1839  
1840  
1841  
1842  
1843  
1844  
1845  
1846  
1847  
1848  
1849  
1850  
1851  
1852  
1853  
1854  
1855  
1856  
1857  
1858  
1859  
1860  
1861  
1862  
1863  
1864  
1865  
1866  
1867  
1868  
1869  
1870  
1871  
1872  
1873  
1874  
1875  
1876  
1877  
1878  
1879  
1880  
1881  
1882  
1883  
1884  
1885  
1886  
1887  
1888  
1889  
1890  
1891  
1892  
1893  
1894  
1895  
1896  
1897  
1898  
1899  
1900  
1901  
1902  
1903  
1904  
1905  
1906  
1907  
1908  
1909  
1910  
1911  
1912  
1913  
1914  
1915  
1916  
1917  
1918  
1919  
1920  
1921  
1922  
1923  
1924  
1925  
1926  
1927  
1928  
1929  
1930  
1931  
1932  
1933  
1934  
1935  
1936  
1937  
1938  
1939  
1940  
1941  
1942  
1943  
1944  
1945  
1946  
1947  
1948  
1949  
1950  
1951  
1952  
1953  
1954  
1955  
1956  
1957  
1958  
1959  
1960  
1961  
1962  
1963  
1964  
1965  
1966  
1967  
1968  
1969  
1970  
1971  
1972  
1973  
1974  
1975  
1976  
1977  
1978  
1979  
1980  
1981  
1982  
1983  
1984  
1985  
1986  
1987  
1988  
1989  
1990  
1991  
1992  
1993  
1994  
1995  
1996  
1997  
1998  
1999  
2000  
2001  
2002  
2003  
2004  
2005  
2006  
2007  
2008  
2009  
2010  
2011  
2012  
2013  
2014  
2015  
2016  
2017  
2018  
2019  
2020  
2021  
2022  
2023  
2024  
2025  
2026  
2027  
2028  
2029  
2030  
2031  
2032  
2033  
2034  
2035  
2036  
2037  
2038  
2039  
2040  
2041  
2042  
2043  
2044  
2045  
2046  
2047  
2048  
2049  
2050  
2051  
2052  
2053  
2054  
2055  
2056  
2057  
2058  
2059  
2060  
2061  
2062  
2063  
2064  
2065  
2066  
2067  
2068  
2069  
2070  
2071  
2072  
2073  
2074  
2075  
2076  
2077  
2078  
2079  
2080  
2081  
2082  
2083  
2084  
2085  
2086  
2087  
2088  
2089  
2090  
2091  
2092  
2093  
2094  
2095  
2096  
2097  
2098  
2099  
2100  
2101  
2102  
2103  
2104  
2105  
2106  
2107  
2108  
2109  
2110  
2111  
2112  
2113  
2114  
2115  
2116  
2117  
2118  
2119  
2120  
2121  
2122  
2123  
2124  
2125  
2126  
2127  
2128  
2129  
2130  
2131  
2132  
2133  
2134  
2135  
2136  
2137  
2138  
2139  
2140  
2141  
2142  
2143  
2144  
2145  
2146  
2147  
2148  
2149  
2150  
2151  
2152  
2153  
2154  
2155  
2156  
2157  
2158  
2159  
2160  
2161  
2162  
2163  
2164  
2165  
2166  
2167  
2168  
2169  
2170  
2171  
2172  
2173  
2174  
2175  
2176  
2177  
2178  
2179  
2180  
2181  
2182  
2183  
2184  
2185  
2186  
218

# FIGURE 224

ATGGCTGGTGAACGGCGGGGCGGGCAGGGGACCGGGGCGGGCCCGGGAGCGGGCCAGCTGCGGGAGCCCTGA  
ATCACCGCTGGCGCCGACTCCACCATGAACCTGCGCTGCAGGAGCTGGGAGCTGGCAGCAACGTGGGATTCCAG  
AAGGGGACACAGCAGCTGTAGGCTCACGCACGCAGCTGGAGCTGGTCTTACAGAGTGGCTCTCTACTGCTGGCT  
GCACTGCTTTCTGGGCTGCTTGTGGCCCTAGGGGTCCAGTACACAGAGACCCATCCACAGCAGCTTGCCTTACA  
GAGGCTGCACTTGCAGTGGCTGGAAAAATCTGGAGTCCCTGGACCGAGGGGTGAGCCCTGTGAGGACTTTTAC  
CAGTCTCTCTGTGGGGCTGGATTTCGGAGGAACCCCTGCCCGATGGGGCTTCTCGTGGAAACACTTCAACAGC  
CTCTGGGACCAAAACAGGCCATATGAAGCAGCTGCTTGAACACACACTTCAACTCCAGCAGTGAAGCTGAG  
CAGAAGACACAGCGCTTCTACCTATCTTGCTTACAGGTGGAGCGCATTGGAGAGCTGGGAGCCAGGCCACTGAGA  
GACCTCATTTGAGAAGATTGGTGGTTGGAACATTACGGGGCCCTGGACCAAGGACAACTTTATGGAGGTGTTGAAG  
GCAGTAGCAGGGACCTACAGGGCCACCCCATTTCTACCGCTCTACATCAGTCCGAGCTCTAAGAGTTCACACAGC  
AATGTTATCCAGGTGGACAGCTCTGGGCTCTTCTGCCCTCTCGGGATTACTACTTAAACAGAACTGCCAATGAG  
AAAGTGCTCACTGCTATCTGGATTACATGGAGGAACCTGGGAGTGCTGCTGGGTGGGCGCCCACTCCACGAGG  
GAGCAGATGCAGCAGGTGCTGGAGTTGGAGATACAGCTGGCCAAACATCACAGTGCCCCAGGACACAGCGCGCGAC  
GAGGAGAAGATCTACCAACAAGTAGCATTTTCGGAGCTGCAGGCTCTGGGCGCCCTCCATGGACTGGCTTGAATTC  
CTGTCTTTCTTGCTGTCAACATTGGAGTTGAGTGACTCTGAGCCCTGTGGTGGTGTATGGGATGGATTATTTGAG  
CAGGTGTTCAGAGCTCATCAACCGCACGGAACCAAGCATCCTGAACAATTAACCTGATCTGGAACCTGGTGAACAAAG  
ACAACCTCAAGCTGGACCGACGCTTTGAGTCTGCACAAGAGAAGCTGTGGGAGACCTCTATGGCACTAAGAAG  
TCCTGTGTGCCGAGGTGGCAGACCTGCATCTCCAACCGGATGAGCGCTTGGCTTTGGCTTTGGGGTCACTCTTC  
GTGAAGGCCACGTTTGACCGCAAGCAAGAAATTCAGAGGGGATGATCAGCGAAATCCGGACCGCATTTGAG  
GAGGCCCTGGGACAGCTGGTTTGGATGGATGAGAAGACCCCGAGGCGAGCCAGGAGAAGACAGATGCCATCTAT  
GATATGATTGGTTTCCGAGCTTTATCCTGGAGCCCAAGAGCTGAGTGATGTTTATGACGGGTGCAAAATTTCT  
GAAGATTCTTTCTCCAACCATGTTGAATTTGTACAACCTCTCTGCCAAGGTTTATGGCTGACCAAGCTCCGCAAG  
CTCCGAGCGGAGACCAAGTGGAGCATGACCCCCAGACAGTGAATGCCCTACTACTTCCAACCTAAGAATGAGATC  
GTCTTCCCGCTGGCATCTCTGACGGGCCCTTCTATGCCCGCAACCAACCCAGGCGCTGAACCTCTCGTGGCATC  
GTTGTGGTCTATGGGCTAGGTTGACGCATGCCCTTGTATGACCAAGGGCGCAGTATGACAAAGAGGGAACCTG  
CGGCCTGGTGGCAGAATGAGTCCCTGGCAGCCTTCCGGAACCAACGCGCTGCTATGGAGGAACAGTACATCA  
TACCAGTCAATGGGAGAGCTCAACGCGCGCCAGACGCTGGGGAGAACTATTACCAACCGGGGCTGAAG  
GCTGCTACATGCTTCAAAAGCATGGCTGAGAAAGCATGGGAGGAGCAGCAACTGCCAGCGTGGGGCTCAC  
AACCACAGCTCTTCTTCTGGGATTGGCCAGGTTGTGGTCTCGGTCCGACACACAGAGAGCTCTACAGAGGG  
CTGGTACCGACCCCAAGCCCTGCCCGCTTCCGCGTGTGGGCACTCTCTCAACTCCCGTGACTTCTTGGG  
CACTTCGCTGCCCTGTGGCTCCCCATGAACCCAGGGCAGCTGTGTGAGGTGTGGTAGACCTGGATCAGGGGA  
GAAATGGGCACTGTCAACAGACTGGGGCAGCTCTCTGACAAAGCTGTTTGTCTTGGGTTGGGAGGAAGCAA  
ATGCAAGCTGGGCTGGGCTAGTCCCTCCCCCAGAGTGACATGAGTACAGACCTCTCTCAATCACACATTG  
TGCCTCTGCTTTGGGGGTGCCCTGCTCCAGCAGAGCCCCACCATTCACTGTGACATCTTCCGTGTCACTCT  
GCTTGAAGAGGTCTGGGTGGGGAGGCCAGTTCCCATAGGAAGAGTCTGCC

1007081.122401

## **FIGURE 225**

MNVALQELGAGSNVGFQKGTQQLLSRTQLELVLAGASLLLAALLLGCLVALGVQYHRDPSH  
STCLTEACIRVAGKILES LDRGVSPCEDFYQFSCGGWIRRNPLDGRSRWNTFNSLWDQNQA  
ILKHLLENTTFNSSSEAEQKTQRFYLSCLQVERIEELGAQPLRLDIEKIGWNITGPWDQDN  
FMEVLKAVAGTYRATPFFTIVYISADSKSSNSNVIQVDQSGFLFLPSRDYYLNRTANKEVLTAY  
LDYMEELGMLLGGRPSTSTREQMQQVLELEIQLANITVPQDQRRDEEKIYHKMSISELQALAP  
SMDWLEFLSFLLSPLELSDSEPVVVGMDYLQQVSELINRTEPSILNNYLIWNLVQKTTSSL  
DRRFESAQEKLLLETLYGTTKKSCVPRWQTCISNTDDALGFALGSLFVKATFDRQSKETAEAGMI  
SEIRTAEEALGQLVWMDEKTRQAAKEKADAIYDMIGFPDFILEPKELDDVYDGYEISEDSF  
FQNMLNLNYNFSKVMADQLRKPPSRDQWSMTPQTVNAYYLPKNEIVFPAGILQAPFFYARNH  
PKALNFGGIGVVMGHELTHAFDDQGREYDKEGNLRPWWQNESLAAFRNHTACMEEQYNQYQV  
NGERLNGRQTLGENTIDNGGLKAAYNAYKAWLRKHGEEQQLPAVGLTNHQLFFVGFQAQVWCS  
VRTPESSHEGLVTDPHSPARFRVLGTLSNSRDFLRHFGCPVGGSPMNPQGQCEVW

**Type II Transmembrane domain:**

amino acids 32-57



## **FIGURE 227**

GGCCGAGCGGGGGTGTCTGCGCGGCGGCCGTGATGGCTGGTGACGCGGGGGCCGGGCAGGGGA  
CCGGGGCCGCGGCCCGGGAGCGGGCCAGCTGCCGGGAGCCCTGAATCACCGCCTGGCCCGAC  
TCCACCATGAACGTGCGCTGCAGGAGCTGGGAGCTGGCAGCAACGTGGGATTCCAGAAGGG  
GACAAGACAGCTGTTAGGCTCACGCACGCAGCTGGAGCTGGTCTTAGCAGGTGCCTCTCTAC  
TGCTGGCTGCACGTGCTTCTGGGCTGCCCTTGTGGCCCTAGGGGTCCAGTACCACAGAGACCCA  
TCCCAACAGCACCTGCCCTTACAGAGGCCCTGCATTGAGTGGCTGGAAAAATCCTGGAGTCCCT  
GGACCGAGGGGTGAGCCCCGTGTGAGGACTTTTACCAGTTCCTCTGTGGGGGCTGGATTGGA  
GGAACCCCTGCCCGATGGGCGTTCTCGCTGGAACACCTTCAACAGCCTCTGGGACCAAAAC  
CAGGCCATACTGAAGCACCTGCTTGAAAAACACCACCTTCAACTCCAGCAGTGAAGCTGAGCA  
GAAGACACAGCGCTTCTACCTATCTTGCCCTACAGGTGGAGCGCATTGAGGAGCTGGGAGCCC  
AGCCACTGAGAGACCTCATTGAGAAGATTGGTGGTTGGAACATTACGGGGCCCTGGGACCAG  
GACAACTTTATGGAGGTGTTGAAGGCAGTAGCAGGGACCTACAGGGCCACCCCATTTCTTAC  
CGTCTACATCAGTGCCGACTCTAAGAGTTC AACAGCAATGTTATCCAGGTGGACCAGTCTG  
GGCTCTTTCTGCCCTCTCGGATTACTACTTAAACAGAACTGCCAATGAGAAAGTAAGGAAC  
ATCTTCCGAACCCCCATCCCTACCCCTGGCTGAGCTGGGCTGATCCCTGTTGACTTTTCCCT  
TTGCCAAGGGTCAGAGCAGGGAAGGTGAGCCTATCCTGTCACTAGTGAACAAACTGCCCT  
CCTTTCTTTCTTTCTTTCTTCTCCTCCCTCCCTCCCTTCTTCCCTTTTCTTCTTCTTCTTCC  
TCTTATTCTTCTAGTAGGTTTCATAGACACCTACTGTGTGCCAGGTCCAGTGGGGGAATTG  
GAGATATAAGTTTCCGAGCCATTGCCACAGGAAGCGTTTCAGTGTGATGGGTTTCATGGACCT  
AGATAGGCTGATAACAAGGCTCAAGAGGGTCTTGAGGATTGAGGAGAGACTTATGGAGCC  
AGCAAAGTCTTCTCTGAAGAGATTGCATTTGAGCCAGGTCTCTGTAG

## FIGURE 228

ATGCCTACTACCTTCCAATAAGAATGAGATCGTCTTCCCCGCTGGCATCCTGCAGGCCCC  
TTCTATGCCCCGCAACCACCCCAAGGCCCTGAACCTCGGTGGCATCGGTGTGGTCATGGGCCA  
TGAGTTGACGCATGCCTTTGATGACCAAGGGCGCGAGTATGACAAAGAAGGGAACCTGCGGC  
CCTGTTGGCAGAAATGAGTCCCTGGCAGCCTTCCGGAACCAACACGGCCTGTCATGGAGGAACAG  
TACAATCAATACCAGGTCAATGGGGAGAGGCTCAACGGCCGCCAGACGCTGGGGGAGAACAT  
TGCTGACAACGGGGGGCTGAAGGCTGCCTACAATGCTTACAAAGCATGGCTGAGAAAGCATG  
GGGAGGAGCAGCAACTGCCAGCCGTGGGGCTCACCAACCACCACTCTTCTTCGTGGGATTT  
GCCCAGGTGTGGTGCTCGGTCCGCACACCAGAGAGCTCTCACGAGGGGCTGGTGACCGACCC  
CCACAGCCCTGCCCGCTTCCGCGTGCTGGGCACTCTCTCCAATCCCGTGACTTCCTGCGGC  
ACTTCGGCTGCCCTGTCCGCTCCCCATGAACCCAGGGCAGCTGTGTGAGGTGTGGTAGACC  
TGGATCAGGGGAGAAATGGCCAGCTGTCCACAGACCTGGGGCAGCTCTCCTGACAAAGCTGT  
TTGCTCTTGGGTGGGAGGAAGCAAATGCAAGCTGGGCTGGGTCTAGTCCCTCCCCCCCCACA  
GGTGACATGAGTACAGACCCTCCTCAATCACCACATTGTGCCTCTGCTTTGGGGGTGCCGCT  
GCCTCCAGCAGAGCCCCACCACTTCACTGTGACATCTTCCGTGTACCCCTGCCTGGAAGAG  
GTCTGGGTGGGGAGGCCAGTTCCCATAGGAAGGAGTCTGCCTCTTCTGTCCCCAGGCTCACT  
CAGCCTGGCGGCCATGGGGCTGCCGTGCCTGCCCCACTGTGACCCACAGGCCTGGGTGGTG  
TACCTCCTGGACTTCTCCCCAGGCTCACTCAGTGCACCTTAGGGGTGGACTCAGCTCTGTCT  
TGGCTCACCCCTCACGGGCTACCCCACTCACCCTGTGCTCCTTGTGCCACTGCTCCCAGTG  
CTGCTGCTGACCTTCACTGACAGCTCCTAGTGAAGCCCAAGGGCCTCTGAAAGCCTCCTGC  
TGCCCACTGTTTCCCTGGGCTGAGAGGGGAAGTGCAATATGTGTAGCGGGTACTGGTTCTGT  
GTCTTAGGGCACAAGCCTTAGCAAATGATTGATTCTCCCTGGACAAAGCAGGAAAGCAGATA  
GAGCAGGGAAGGAACAGAGTTTATTTTACAGAAAAGAGGGTGGGAGGGTGTGGTCT  
TGGCCCTTATAGGACC

10017087.102444

[illegible]

CCACACGGGTCCGAGCGCCCGAGAATTAGACACACTCCGGACGCGGCCAAAAGCAACCAGGAA  
GGAGGGGAGAAACAAACACCGAAAAACAAAAGAGAGAGAAACAAACACCCCAACTCGGGGTGG  
GGGGAAGAAAGAAAAGAAAGAAACACCCACCCACCAAAAAAAAATAAAAAAAA  
AAAAAAAATAATCTGTGTGGCGCGCGCTGTGTTCCCGGGAAGACTCGCCAGCACCGGGGG  
TGGGGAGGTGCGAGCTGAAAGCTGCTGGAGGTGAGCAGCCTTAGCAGGATGGACAATGAT  
TGTTTGGTGACGGGTGCTTTGTTGCTCGAACAGTGGCTGCGCGGGTGTCTCTCAGCCTGTG  
CTGCCCTGCTACCCCTCTGCTCTCCGGCTGGACAGAGTGTGGACTTCCCTTGGGCGGCCGTGG  
ACAACTATGATGTGTCAGAAAGGGACACGCGCGTGTAGTGTGTTATTTGGAGAGATGGAGCT  
TCAAAGGGTGCTTGGCTGAACCGGTCAAGTATATTTTTGCGGAGGTTGATAGTGTGCTAGT  
GGATCTCTCGATGTTCAATTTCAACATTGAAATAAAGGGACACAGCCTCCAGATACAGAATG  
TAGATGTGACAGATGATGCCCCATACAGCTGTTTCTGTTAGACTCAACATACACCGAACA  
ATGCAAGTGCATCTAATGTGCAAGTTCCTCCTAAGATATATGACATCTCAAATGATATGAC  
CGTCAATGAAGGAACCAACGCTCACTCTTACTTGTTTGGCCACTGGGAAACAGAGGCTTCCA  
TTTTCTTGGGACACATCTCCCCATGACAAAACATTTGAAATGGAGACATATTGGACATT  
TATGGAATTGACAGGGACAGGCTGGGGAATATGAATTCAGTGGCGGAAATGCTGTGTCAAT  
CCGAGATTGAGGAAAGTAAAAAGTTGTGTCAACTTGTCTCTACTATTGAGAAATTAAT  
CTGGCACCGTGACCCCGGACGCGAGTGGCTGATAGATGTGAAAGTGGAGGTGCGGCTGCGCCT  
CCAGCCTTTGAATGTGACAAAGGAGGAGAGAAGCTCTTCAATGGCCAAACAGGAATATTAT  
TCAAAATTTTAGCACAGATCCATTCTACTGTGTACCAAGTGCACAGGACACTTCGGCA  
ATTATACCTGTGTGGCTGCCAACAGCTAGGCACAAACCAATGCGAGCCTGCCTCTTAACCCCT  
CCAAGTACAGCCGAGATGTGAATTACCGGGAGCGCTGATGTTCTTTTCTGCTGGTACTCT  
TGTGTGACAGTGCTCTCTTTACACAGATGTTACTCTGAGAAGATGCCATTCTACAAATAA  
TCAAGAGGCTTCAATAAAGGCTTTTAAGGATTTCTGAAAGTCTGATGGCTGGATCCAATCT  
GGTCAAGTTTGTAAAAAGCAGCGTGGGATATAATCAGCAGCTGTTTACATGGGGATGATCGCC  
TTCTGTAGAATTGCTCATTATGTAAATACTTTAACTTCTACTCTTTTGTATAGTCATATTA  
CCTTGTGAAGCAGTACACATTTGCTTTTTTAAAGACGTGAAAGCTGTAATTTACTTTTAA  
AGGATATTAATGTGATTTCTATGTTTGAATCTACAACTTTTCAAAGCATTCAGTCATGTG  
CTGCTAGGTTGCAAGCTGTAGTTTACAAAAACGAATATTGCGAGTGAATTTGTGATTTCTTAA  
GGCTGCAATACAAGCATTCAGTTCCCTGTTTTCATTAAGATGACCATCCATCTTACAAGATG  
CATTTTTTTCTTTTGTATAAAAAAGCAATAATTGGCTTCAGATTAATTTCTCAAAATA  
TAACACATCTAGATTTTCTGCTTGATGATTAATAGGTTTCAGGAATGAGCCTTGTAAT  
ATAACTGGCTGTGCACTGCTCTCTTTCTTCTGTAAGTTTTCAGCATGGGTGTGCTCTCATAC  
AATAATATTTTTCTCTTTGCTTCAACTAATATAAAATGTTTTGCTAAATCTTCAAAATTTGA  
AAGTAAAAATAAACAGAGTATCAGTTTAAACCATACATCTCTCTAAGTACGGAAGAGC  
TATTGGACGTGAAAAATCTCTTCTCGCAGTGCACATGGGTTTGAATTTTGGCCCACT  
AATCAGTTTCTGTGATGAGAGCAAAATTTAATAACAGTATAGTAAATATACCATATGATTTCT  
TTTAGTTGTAGCTAAATGTTAGATCAACCGTGGGAATATCTCCCTTTAAATATGACAGACA  
GTCCACTCAAAGGATGCCTAGCAATACAGCATCTTTTTCTTTCTAGTGTGCAAGCAAAAA  
TTTTAAGATGATTTGTGACAAAGGGGACAAAGTCCATACCACTAATTTACAAGATTTGGTA  
AGCGCTCATCATTAATTTATTTTGTGTCAGGTTATATGACAGCTGCACCTGGAGGGTATGGA  
TATGGATATGGACGTTCCAGAGACTATAATGGCAGAAACAGGTTGGTTATAGCCGCTATCTC  
AGGAGAAATATACGACACAAATATGACAACTGAAATGAGACATGCACATTAATATAGATACA  
CAAGGAATAATTTCTGATGACAGGATGCTCCTTCAAATGGCTGATATTATAAAGGTTTTGG  
AGCTGCACTGAAGCATCTTATTTTATAGTATATCAACCTTTTGTTTTTAAATTTAGCTGCCA  
AGGTAGCTGAAGACCTTTTAGACAGTCCATCTTTTTTTTTAAATTTTTCTGCTATTTAA  
AGACAAATTTAGGGACGTTTTGTCAAAAAAAAATAAAAAAAAATAAAAAAAA

## **FIGURE 230**

MLLVQGACCSNQWLA AVL LLSLCCLLPSCLPAGQSVDFPWA AVDNMMVRKGD TAVLR CYLED  
GASKGAWLNRSSII FAGGDKWSVDPRVSI STL NKR DYSLQIQNV DVTDDGPYTC SVQTQHTP  
RTMQVHLTVQVPPKIYDISNDMTVNEGTVNLTLCLATGKPEPSISWRHISPSAKPFENGQYL  
DIYGITRDQAGEYECSAENAVSFDPVRKV KVVVNFAPT IQEIKSGTVTPGRSGLIRCEGAGV  
PPPAFEWYKGEKKLFNGQQGII IQNFSTRSILTVTNVTQEHFGNYTCVAANKLGTTNASLPL  
NPPSTAQYGITGSADVLFSWCWYLVLTLSSTSI FYLKNAIQ

### **Important features of the protein:**

#### **Signal peptide:**

amino acids 1-31

#### **Transmembrane domain:**

amino acids 326-345

#### **N-glycosylation sites.**

amino acids 71-75, 153-157, 273-277, 284-288, 292-296, 305-309

#### **Casein kinase II phosphorylation site.**

amino acids 147-151, 208-212, 224-228

#### **Tyrosine kinase phosphorylation site.**

amino acids 178-186

#### **N-myristoylation sites.**

amino acids 7-13, 63-70, 67-73, 151-157, 239-245, 291-297,  
302-308, 319-325

#### **Myelin P0 protein:**

amino acids 92-121

## FIGURE 231

AGTGGTTCGATGGGAAGGATCTTTCTCCAAGTGGTTCCTCTTGAGGGGAGCATTTCTGCTGG  
CTCCAGGACTTTGGCCATCTATAAAGCTTGGCAATGAGAAAATAAGAAAATTCTCAAGGAGGA  
CGAGCTCTTGAGTGAGACCCAAAGCTGCTTTTACCACAAATTGCAATGGAGCCTTTTCGAAA  
TCAATGTTCCAAAGCCCAAGAGGAGAAAATGGGGTGAACCTTCTCCCTAGCTGTGGTGGTCATC  
TACCTGATCCTGCTCACCCTGGCGCTGGGCTGCTGGTGGTCCAGATTCTGAATCTGCAGGC  
GCGGCTCCGGGTCTGGAGATGTATTTCTCAATGACACTCTGGCGGCTGAGGACAGCCCCT  
CCTTCTCCTTGCTGCAGTCAGCACACCTGGAGAACCTGGCTCAGGTTGCATCGAGGCTG  
CAAGTCTGCAGGCCCAACTCACCTGGGTCCCGCTCAGCCATGAGCACTTGCTGCAGCGGGT  
AGACAACTTCACTCAGAACCCAGGGATGTTCAGAATCAAAGGTGAACAAGGCGCCCCAGGTC  
TTCAAGGTCAAGGGGGCCATGGGCATGCCCTGGTGGCCCTGGCCCGCCGGGACCACTTGCT  
GAGAAGGGAGCCAAGGGGGCTATGGGACGAGATGGAGCAACAGGCCCTCGGGACCCCAAGG  
CCCACCGGAGTCAAGGGAGAGGCGGGCTTCCAGGACCCAGGGTGTCCAGGGAAGCAAG  
GAGCCACTGGCACCCAGGACCCCAAGGAGAGAAGGGCAGCAAGGCCATGGGGGTCTCATT  
GGCCCAAAAGGGGAACTGGAACCTAAGGGAGAGAAAAGGAGACCTGGGTCTCCAGGAAGCAA  
AGGGGACAGGGGCATGAAAGGAGATGCAGGGGTCTAGGGGCTCTGGAGCCAGGGGAGTA  
AAGGTGACTTCGGGAGGCCAGGCCACCAGGTTTGGCTGGTTTCTGGAGCTAAAGGAGAT  
CAAGGACAACCTGGACTGCAGGGTGTTCGGGCCCTCTGGTGCAGTGGGACACCCAGGTGC  
CAAGGGTGAGCCTGGCAGTGCTGGCTCCCCTGGGCGAGCAGGACTTCCAGGAGCCCCGGGA  
GTCCAGGAGCCACAGGCCTGAAAGGAAGCAAAGGGGACACAGGACTTCAAGGACAGCAAGGA  
AGAAAAGGAGAATCAGGAGTTCCAGGCCCTGCAGGTGTGAAGGGAGAACAGGGGAGCCAGG  
GCTGGCAGTCCCAAGGGAGCCCCTGGACAAGCTGGCCAGAAGGGAGACCAGGGAGTGAAAG  
GATCTTCTGGGGAGCAAGGAGTAAAGGGAGAAAAAGGTGAAAGAGGTGAAAACTCAGTGCTC  
GTCAGGATTGTGGCAGTAGTAACCGAGGCCGGGCTGAAGTTTACTACAGTGGTACCTGGGG  
GACAAATTTGCGATGACGAGTGGCAAAATCTGATGCCATTGTCTTCTGCCCATGCTGGGTT  
ACTCCAAGGAAGGGCCCTGTACAAAGTGAGGCTGGCACTGGGCAGATCTGGCTGGATAAT  
GTTAGTGTTCGGGGCAGGAGAGTACCTGTGGAGCTGCACCAAGAATAGCTGGGGCCATCA  
TGACTGCAGCCACGAGGAGGACGACGGCGTGGAGTGACAGCGTCTGACCCGGAACCTTTTCA  
CTTCTCTGCTCCCGAGGTGTCTCGGGCTCATATGTGGGAAGGCAGAGGATCTCTGAGGAGT  
TCCCTGGGGACAACCTGAGCAGCCTCTGGAGAGGGGCCATTAATAAAGCTCAACATCATTTGA

## FIGURE 232

></usr/seqdb2/sst/DNA/Dnaseqs.full/ss.DNA68886  
><subunit 1 of 1, 520 aa, 1 stop  
><MW: 52658, pI: 9.16, NX(S/T): 3  
MRNKKILKEDELLSETQQAAPHQIAMEPFEINVPKPKRRNGVNFSLAVVVIYLILLTAGAGL  
LVVQVLNLQARLRVLEMYFLNDTLAEDSPSFSLLQSAHPGEHLAQGASRLQVLQAQLTWVR  
VSHEHLLQRVDNFTQNPMFRIKGEQGAPGLQGHKGAMGMPGAPGPPGPPAEKGAKGAMGRD  
GATGSPGPPQPPGVKGEAGLQQPQGAPGKQGATGTPGPQGEKGSKGDGGLIGPKGETGTKGE  
KGDLLGLPGSKGDRGMKGDAGVMGPPGAQGSKGDGFRPGPPGLAGFPGAAGKGDQGGPGLQGVPC  
PPGAVGHPGAKGEPGSAGSPGRAGLPGSPGSPGATGLKGSKGDGTLQGGQGRKGESGVPGPA  
GVKGEQGSPLAGPKGAPGQAGQKGDQGVKGSSEGEQGVKGEKGERGENSVSVRIVGSNNRGR  
AEVYYSGTWGTICDDEWQNSDAIVFCRMLGYSKGRALYKVGAGTGQIWLNDNVQCRGTSTLW  
SCTKNSWGHHDSCSHEEDAGVECSV

### **Transmembrane domain:**

amino acids 47-66 (type II)

### **N-glycosylation sites.**

amino acids 43-47, 83-87, 136-140

### **Tyrosine kinase phosphorylation site.**

amino acids 432-440

### **N-myristoylation sites.**

amino acids 41-47, 178-184, 253-259, 274-280, 340-346, 346-352,  
400-406, 441-447, 475-481, 490-496, 515-521

### **Amidation site.**

amino acids 360-364

### **Leucine zipper pattern.**

amino acids 56-78

### **Speract receptor repeat**

amino acids 422-471, 488-519

### **Clq domain proteins.**

amino acids 151-184, 301-334, 316-349

## FIGURE 233

CCCACGCGTCCGAAGGCAGACAAAGGTTTCATTGTAAAGAAGCTCCTTCCAGCACCTCCTCT  
CTTCTCCTTTTGCCCAAACCTCACCCAGTGAGTGTGAGCATTTAAGAAGCATCCTCTGCCAAG  
ACCAAAGGAAAGAAGAAAAAGGGCCAAAAGCCAAATGAAACTGATGGTACTTGTTTTCAC  
CATTTGGGCTAACTTTGCTGCTAGGAGTTCAAGCCATGCCTGCAAATCGCCTCTCTTGCTACA  
GAAAGATACTAAAGATCACAACCTGTACAACCTTCCGGAAGGAGTAGCTGACCTGACACAG  
ATTGATGTCAATGTCCAGGATCATTTCTGGGATGGGAAGGGATGTGAGATGATCTGTTACTG  
CAACTTCAGCGAATTGCTCTGCTGCCAAAAGACGTTTCTTTGGACCAAAGATCTCTTTG  
TGATTCTTGCAACAATCAATGAAGAATCTTCATGTATTCTGGAGAACACCATTCTGATTTC  
CCACAAACTGCACTACATCAGTATAACTGCATTTCTAGTTTCTATATAGTGCAATAGAGCAT  
AGATTCTATAAATTTCTTACTTGTCTAAGACAAGTAAATCTGTGTTAAACAAGTAGTAATAAA  
AGTTAATTCAATCTAAAAAAAAAAAAA

10017001-10017001

## FIGURE 234

</usr/seqdb2/sst/DNA/Dnaseqs.min/ss.DNA52758

<subunit 1 of 1, 98 aa, 1 stop

<MW: 11081, pI: 6.68, NX(S/T): 1

MKLMVLVFTIGLTLLLGVSQAMPANRLSCYRKILKDHCHNLPEGVADLTQIDVNVQDHFWDG  
KGCEMICYCNFSELLCCPKDVFFGPKISFVIPCNNQ

**Important features:**

**Signal peptide:**

amino acids 1-20

**N-glycosylation site.**

amino acids 72-76

**Tyrosine kinase phosphorylation site.**

amino acids 63-71

FIGURE 234

# FIGURE 235

CCCACGCGTCCGCGGACGCGTGGGCTGGACCCAGGCTCTGGAGCGAATTCAGCCTGCAGGG  
 CTGATAAGCGAGGCATTAGTGAGATTGAGAGAGACTTTACCCGCGCGTGGTGGTGGAGGGC  
 GCGCAGTAGAGCAGCAGCAGCGCGGGTCCGCGGAGGCGGCTCTGCTCGCGCGGAGATG  
 TGGAAATCTCCTTCACGAAACGACTCGGCTGTGGCCACCGCGCGCCCGCGCTGGCTGTG  
 CGCTGGGCGCTGGTGCTGGCGGGTGGCTTCTTCTCCTCGGCTTCTCTTCGGGTGGTTTA  
 TAAAAATCCTCCAATGAAGCTACTAACATTACTCCAAAGCATAATATGAAAGCATTTTTGGAT  
 GAATTGAAAGCTGAGAACATCAAGAAGTTCTTACATAATTTTACACAGATACCACATTTAGC  
 AGGAACAGAACAAAACCTTCAGCTTGCAAAGCAAATTCAAATCCGATGGAAAGAATTTGGCC  
 TGGATTCTGTTGAGCTAGCTCATTATGATGTCCTGTTGTCTTACCCAAATAAGACTCATCCC  
 AACTACATCTCAATAATTAAATGAAGATGGAAATGAGATTTCACACATCATTTATTGAACC  
 ACCTCCTCCAGGATATGAAAATGTTTCGGATATTGTACCACCTTCAGTGCTTTCTCTCCTC  
 AAGGAATGCCAGAGGGCGATCTAGTGTATGTTAACTATGCACGAACTGAAGACTTCTTTAAA  
 TTGGAACGGGACATGAAAATCAATTGCTCTGGGAAAATTTGAATTCAGAGATATGGGAAAGT  
 TTTTCAGAGGAAATAAGGTTAAAAATGCCAGCTGGCAGGGGCCAAAGGAGTCATTCTCTACT  
 CCGACCCCTGCTGACTACTTTGCTCCTGGGGTGAAGTCCCTATCCAGACGGTTGGAATCTTCT  
 GGAGTGTGTGTCAGCGTGGAATAATCCTAAATCTGAATGTCAGGAGACCTCTCACACC  
 AGGTTACCCAGCAAATGAATATGCTTATAGGCGTGGAAATTCAGAGGCTGTTGGTCTTCCAA  
 GTATTCCTGTTTCATCCAATTGGATACTATGATGCACAGAGCTCCTAGAAAAAATGGGTGGC  
 TCAGCACCACAGATAGCAGCTGGAGAGGAAGTCTCAAAGTGCCCTCAATGTTGGACCTGG  
 CTTTACTGGAACCTTTTCTACAAAAAGTCAAGATGCACATCCACTCTACCAATGAAGTGA  
 CGAGAATTACAATGTGATAGGTACTCTCAGAGGAGCAGTGGAAACAGACAGATATGTCATT  
 CTGGGAGGTCAACGGGACTCATGGGTGTTTGGTGGTATTGACCTCAGAGTGGAGCAGCTGT  
 TGTTTCATGAAATTGTGAGGAGCTTTGGAACTGAAAAAGGAAGGTTGGAGACCTTAGAAGAA  
 CAATTTTGTTTGCAAGCTGGGATGCAGAAGAATTTGGTCTTCTTGGTTACTCTGAGTGGGCA  
 GAGGAGAATTCAAGACTCCTTCAAGAGCGTGGCGTGGCTTATATTAATGCTGACTCATCTAT  
 AGAAGGAAACTACACTCTGAGAGTTGATTGTACACCGCTGATGTACAGCTTGGTACACAACC  
 TAACAAAAGAGCTGAAAAGCCCTGATGAAGGCTTTGAAGGCAAATCTCTTATGAAAGTTGG  
 ACTAAAAAAGTCCCTCCCGAGATTGAGTGGCATGCCAGGATAAGCAAATTTGGGATCTGG  
 AAATGATTTTGGAGTGTCTTCCAAAGCTTGGAAATGCTTCAAGCAGAGCAGCGGTATACTA  
 AAAATTGGGAAACAAACAAATTGAGCGGCTATCCACTGTATCACAGTGTCTATGAAACATAT  
 GAGTTGGTGGAAAAGTTTATGATCCAATGTTTAAATATCAACCTCATGTTGGGCCAGGTTTCG  
 AGGAGGGATGGTGTGAGCTAGCCAAATCCATAGTGCTCCCTTTTGGATTGTGAGATTATG  
 CTGTAGTTTTAAGAAAGTATGCTGACAAAATCTACAGTATTTCTATGAAACATCCACAGGAA  
 ATGAAGACATACAGTGTATCATTTGATTCACTTTTTCTGTCAGTAAAGAATTTTACAGAAAT  
 TGCTTCCAAGTTCAGTGAGAGACTCCAGGACTTTGACAAAAGCAACCCAAATGATATTAAGAA  
 TGATGAATGATCAACTCATGTTTCTGGAAGAGCATTATTGATCCATTAGGGTTACAGAGC  
 AGGCTCTTTTATAGGCATGTCTATGCTTCAAGCAGCCACAACAGTATGCAGGGGAGTC  
 ATTTCCAGGAATTTATGATGCTCTGTTTGATATTGAAAGCAAAGTGGACCTTCCAAGGCTT  
 GGGGAGAGTGAAGAGACAGATTATGTTGACGCTTCAGAGTCAGGCAGCTGCAGAGACT  
 TTGATGAAGTAGCCTTAAGAGGATTTTATAGAAATCCGATATGAATTTGTTGGTATGTGCTCA  
 CTGAGAAAGATCGTAAATGGGTATATTGATAAATTTAAAATTGGTATATTGAAATAAAGT  
 TGAATATTATATATA

## FIGURE 236

></usr/seqdb2/sst/DNA/Dnaseqs.full/ss.DNA52756

><subunit 1 of 1, 750 aa, 1 stop

><MW: 84305, pI: 6.93, NX(S/T): 10

MWNLHETDSAVATARRPWL CAGALVL AGGFLLGLFGWFIKSSNEATNITPKHNMKAF L  
DELKAENIKKFLHNFTQIPHLAGTEQNFQLAKQIQSQWKEFGLDSVELAHYDVLLSYPNKTH  
PNYISIINEDGNEIFNTSLFEP PPPGYENVSDIVPPPSAFSPQGMPEGDLVYVNYARTEDFF  
KLERDMKINCSGKIV IARYGKVFGRGNKVNAQLAGAKGVILYSDPADYFAPGVKSYPDGWN L  
PGGGVQQRGNILNLNGAGDPLTPGYPAN EYAYRRGIAEAVGLPSIPVHP I GYYDAQKLEKMG  
GSAPPDSSWRGSLKVPYNVGP GFTGNFSTQKV KMHISTNEVTRIYNVIGTLRGAVEPD RYV  
ILGGHRDSWVFGGIDPQSGAAV VHEIVRSFGTLKKEGWRPRRTILFASWDAEEFGLLGSTEW  
AEENSRLQLQERG VAYINADSSIEGNYTLRV DCTPLMYSLVHNLT KELKSPDEGFEGKSLYES  
WTKKSPSP EFSGMPRI SKLGS GNDFEVFFQRLGIASGRARYTKNWETNKFSGYPLYHSVYET  
YELVEKFYDPMFKYHLTVAQVRGGMVFELANSIVLPFD CRDYAVVLRKYADKIYSISM KHPQ  
EMKTYSVS FDSLFS AVKNFTEIASKF SERLQDFDKSNPIVLRMMNDQLMFLERAFIDPLGLP  
DRPFYRHVIYAPSSH NKYAGESFPGIYDALFDIESKVDP SKAWGEVKRQIYVAAFTVQAAAE  
TLSEVA

### Signal sequence:

amino acids 1-40

### N-glycosylation sites.

amino acids 76-80, 121-125, 140-144, 153-157, 195-199, 336-340,  
459-463, 476-480, 638-642

### Tyrosine kinase phosphorylation sites.

amino acids 363-372, 605-613, 606-613, 617-626

### N-myristoylation sites.

amino acids 85-91, 168-174, 252-258, 256-262, 282-288, 335-341,  
360-366, 427-433, 529-535, 707-713

Superfect (Qiagen) and pulse-labeled for 3 hours with [<sup>35</sup>S]methionine and [<sup>35</sup>C]cysteine. Both epitope-tagged proteins co-migrate when 20 microliters of 15-fold concentrated serum-free conditioned medium were electrophoresed on a polyacrylamide gel (Novex) in sodium dodecyl sulfate sample buffer (SDS-PAGE). The VEGF-E-IgG expression plasmid was constructed by cloning the ORF in front of the human Fc (IgG) sequence.

The VEGF-E-IgG plasmid was co-transfected with Baculogold Baculovirus DNA (Pharmlingen) using Lipofectin (GibcoBRL) into 10<sup>5</sup> Sf9 cells grown in Hink's TNM-FH medium (JRH Biosciences) supplemented with 10% fetal bovine serum. Cells were incubated for 5 days at 28°C. The supernatant was harvested and subsequently used for the first viral amplification by infecting Sf9 cells at an approximate multiplicity of infection (MOI) of 10. Cells were incubated for 3 days, then supernatant harvested, and expression of the recombinant plasmid determined by binding of 1 ml of supernatant to 30 µl of Protein-A Sepharose CL-4B beads (Pharmacia) followed by subsequent SDS-PAGE analysis. The first amplification supernatant was used to infect a 500 ml spinner culture of Sf9 cells grown in ESF-921 medium (Expression Systems LLC) at an approximate MOI of 0.1. Cells were treated as above, except harvested supernatant was sterile filtered. Specific protein was purified by binding to Protein-A Sepharose 4 Fast Flow (Pharmacia) column.

#### EXAMPLE 86: Northern Blot Analyses for PRO200

Blots of human poly(A)<sup>+</sup> RNA from multiple adult and fetal tissues and tumor cell lines were obtained from Clontech (Palo Alto, CA). Hybridization was carried out using <sup>32</sup>P-labeled probes containing the entire coding region and washed in 0.1 x SSC, 0.1% SDS at 63°C.

VEGF-E mRNA was detectable in fetal lung, kidney, brain, liver and adult heart, placenta, liver, skeletal muscle, kidney, and pancreas. VEGF-E mRNA was also found in A549 lung adenocarcinoma and HeLa cervical adenocarcinoma cell lines.

#### EXAMPLE 87: In Situ Hybridization of Human Fetal Tissue Sections for PRO200

Formalin-fixed, paraffin-embedded human fetal brain, liver, lower limb, small intestine, thyroid, lymph node, thymus, stomach, trachea, skin, spleen, spinal cord, adrenal, placenta, cord, and adult liver, pancreas, lung, spleen, lymph node, adrenal, heart, aorta, and skin were sectioned, deparaffinized, deproteinized in proteinase K (20 µg/ml) for 15 minutes at 37°C, and further processed for in situ hybridization as described by Lu LH and Gillett NA (Cell Vision 1:169-176, 1994). A [ $\alpha$ -<sup>32</sup>P]UTP-labeled antisense riboprobe was generated from a PCR product of 980 bp (primers GGC GGAATCCAACCTGAGTAG and GCGGCTATCCTCTGTGCTC, SEQ ID NOS: 493 and 494, respectively). The slides were dipped in Kodak NTB2 nuclear track emulsion and exposed for 4 weeks.

VEGF-E mRNA expression included localization at the growth plate region and embracing fetal myocytes.

#### EXAMPLE 88: Myocyte Hypertrophy Assay for PRO200

Myocytes from neonatal Harlan Sprague Dawley rat heart ventricle (23 days gestation) were plated in duplicate at 75000 cells/ml in a 96-well plate. Cells were treated for 48h with 2000, 200, 20, or 2 ng/ml VEGF-E-IgG. Myocytes were stained with crystal violet to visualize morphology and scored on a scale of 3 to

7, 3 being nonstimulated and 7 being full-blown hypertrophy.

2000 ng/ml and 200 ng/ml VEGF-E caused hypertrophy, scored as a 5.

#### EXAMPLE 89: Cell Proliferation Assay for PRO200

Mouse embryonic fibroblast C3H10T1/2 cells (ATCC) were grown in 50:50 Ham's F-12: low glucose DMEM medium containing 10% fetal calf serum (FCS). Cells were plated in duplicate in a 24-well plate at 1000, 2000, and 4000 cells/well. After 48 hours, cells were switched to medium containing 2% FCS and were incubated for 72 hours with 200, 800, or 2000 ng/ml VEGF-E or no growth factor added.

Approximately 1.5 fold greater number of cells were measured in the presence of 200 ng/ml VEGF-E as in its absence, at all three cell densities.

#### EXAMPLE 90: Endothelial Cell Survival Assay for PRO200

Human umbilical vein endothelial cells (HUVEC, Cell Systems) were maintained in Complete Media (Cell Systems) and plated in triplicate in serum-free medium (Basic Media from Cell Systems containing 0.1% BSA) at 20,000 cells/well of a 48-well plate. Cells were incubated for 5 days with 200 or 400 ng/ml VEGF-E-IgG, 100 ng/ml VEGF, 20 ng/ml basic FGF, or no addition.

Survival was 2-3 times greater with VEGF-E as compared to lack of growth factor addition. VEGF and basic FGF were included as positive controls.

#### EXAMPLE 91: Isolation of cDNA Clones Encoding Human PRO285

A proprietary expressed sequence tag (EST) DNA database (LIFESEQ™, Incyte Pharmaceuticals, Palo Alto, CA) was searched and an EST (#2243209) was identified which showed homology to the *Drosophila* Toll protein.

Based on the EST, a pair of PCR primers (forward and reverse):

TAAAGACCCAGCTGTGACCG (SEQ ID NO:499)

ATCCATGAGCCTCTGATGGG (SEQ ID NO: 500), and

a probe:

ATTATGTCTCGAGGAAAGGGACTGGTTACCAGGGCAGCCAGTTC (SEQ ID NO: 501)

were synthesized.

mRNA for construction of the cDNA libraries was isolated from human placenta tissue. The cDNA libraries used to isolate the cDNA clones were constructed by standard methods using commercially available reagents such as those from Invitrogen, San Diego, CA (Fast Track 2). The cDNA was primed with oligo dT containing a NotI site, linked with blunt to SalI hemikinased adaptors, cleaved with NotI, sized appropriately by gel electrophoresis, and cloned in a defined orientation into the cloning vector pCR2.1 (Invitrogen, Inc.) using reagents and protocols from Life Technologies, Gaithersburg, MD (Super Script Plasmid System). The double stranded cDNA was sized to greater than 1000 bp and the cDNA was cloned into BamHI/NotI cleaved vector. pCR2.1 is a commercially available plasmid, designed for easy cloning of PCR fragments, that carries AmpR and KanR genes for selection, and LacZ gene for blue-white selection.

In order to screen several libraries for a source of a full-length clone, DNA from the libraries was screened by PCR amplification with the PCR primer pair identified above. A positive library was then used to isolate clones encoding the PRO285 gene using the probe oligonucleotide and one of the PCR primers.

A cDNA clone was sequenced in entirety. The entire nucleotide sequence of DNA40021-1154 (encoding PRO285) is shown in Figure 208 (SEQ ID NO:495). Clone DNA40021-1154 contains a single open reading frame with an apparent translational initiation site at nucleotide positions 61-63 (Figure 208). The predicted polypeptide precursor is 1049 amino acids long, including a putative signal peptide at amino acid positions 1-29, a putative transmembrane domain between amino acid positions 837-860, and a leucine zipper pattern at amino acid positions 132-153 and 704-725, respectively. It is noted that the indicated boundaries are approximate, and the actual limits of the indicated regions might differ by a few amino acids. Clone DNA40021-1154 has been deposited with ATCC (designation: DNA40021-1154) and is assigned ATCC deposit no.209389.

Based on a BLAST and FastA sequence alignment analysis (using the ALIGN computer program) of the full-length sequence is a human analogue of the *Drosophila* Toll protein, and is homologous to the following human Toll proteins: Toll1 (DNAX# HSU88540-1, which is identical with the random sequenced full-length cDNA #HUMRSC786-1); Toll2 (DNAX# HSU88878-1); Toll3 (DNAX# HSU88879-1); and Toll4 (DNAX# HSU88880-1).

#### EXAMPLE 92: Isolation of cDNA Clones Encoding Human PRO286

A proprietary expressed sequence tag (EST) DNA database (LIFESEQ™, Incyte Pharmaceuticals, Palo Alto, CA) was searched and an EST (#694401) was identified which showed homology to the *Drosophila* Toll protein.

Based on the EST, a pair of PCR primers (forward and reverse):

GCCGAGACAAAACGTTCTCC (SEQ ID NO:502)

CATCCATGTTCTCATCCATTAGCC (SEQ ID NO: 503), and

a probe:

TCGACAACCTCATGCAGAGCATCAACCAAAGCAAGAAAACAGTATT (SEQ ID NO: 504)

were synthesized.

mRNA for construction of the cDNA libraries was isolated from human placenta tissue. This RNA was used to generate an oligo dT primed cDNA library in the vector pRKSD using reagents and protocols from Life Technologies, Gaithersburg, MD (Super Script Plasmid System). pRKSD is a cloning vector that has an sp6 transcription initiation site followed by an SfiI restriction enzyme site preceding the XhoI/NotI cDNA cloning sites. The cDNA was primed with oligo dT containing a NotI site, linked with blunt to SalI hemikinased adaptors, cleaved with NotI, sized to greater than 1000 bp appropriately by gel electrophoresis, and cloned in a defined orientation into XhoI/NotI-cleaved pRKSD.

In order to screen several libraries for a source of a full-length clone, DNA from the libraries was screened by PCR amplification with the PCR primer pair identified above. A positive library was then used to isolate clones encoding the PRO286 gene using the probe oligonucleotide identified above and one of the PCR

primers.

A cDNA clone was sequenced in entirety. The entire nucleotide sequence of DNA42663-1154 (encoding PRO286) is shown in Figure 210 (SEQ ID NO:497). Clone DNA42663-1154 contains a single open reading frame with an apparent translational initiation site at nucleotide positions 57-59 (Figure 211). The predicted polypeptide precursor is 1041 amino acids long, including a putative signal peptide at amino acid positions 1-26, a potential transmembrane domain at amino acid positions 826-848, and leucine zipper patterns at amino acids 130-151, 206-227, 662-684, 669-690 and 693-614, respectively. It is noted that the indicated boundaries are approximate, and the actual limits of the indicated regions might differ by a few amino acids. Clone DNA42663-1154 has been deposited with ATCC (designation: DNA42663-1154) and is assigned ATCC deposit no. 209386.

Based on a BLAST and FastA sequence alignment analysis (using the ALIGN computer program) of the full-length sequence of PRO286, it is a human analogue of the *Drosophila* Toll protein, and is homologous to the following human Toll proteins: Toll1 (DNAX# HSU88540-1, which is identical with the random sequenced full-length cDNA #HUMRSC786-1); Toll2 (DNAX# HSU88878-1); Toll3 (DNAX# HSU88879-1); and Toll4 (DNAX# HSU88880-1).

#### EXAMPLE 93: NF- $\kappa$ B Assay for PRO285 and PRO286

As the Toll proteins signal through the NF- $\kappa$ B pathway, their biological activity can be tested in an NF- $\kappa$ B assay. In this assay Jurkat cells are transiently transfected using Lipofectamine reagent (Gibco BRL) according to the manufacturer's instructions. 1  $\mu$ g pB2XLuc plasmid, containing NF- $\kappa$ B-driven luciferase gene, is cotransfected with 1  $\mu$ g pSR $\alpha$ N expression vector with or without the insert encoding PRO285 or PRO286. For a positive control, cells are treated with PMA (phorbol myristyl acetate; 20 ng/ml) and PHA (phytohaemagglutinin, 2  $\mu$ g/ml) for three to four hours. Cells are lysed 2 or 3 days later for measurement of luciferase activity using reagents from Promega.

#### EXAMPLE 94: Isolation of cDNA Clones Encoding Human PRO213-1, PRO1330 and PRO1449

A consensus DNA sequence was assembled relative to other EST sequences using phrap as described in Example 1 above. This consensus sequence is herein designated DNA28735. Based on the DNA28735 consensus sequence, oligonucleotides were synthesized: 1) to identify by PCR a cDNA library that contained the sequence of interest, and 2) for use as probes to isolate a clone of the full-length coding sequence for PRO213-1, PRO1330 and/or PRO1449. A pair of PCR primers (forward and reverse) were synthesized:

forward PCR primer 5'-TGGAGCAGCAATATGCCAGCC-3' (SEQ ID NO:511)

reverse PCR primer 5'-TTTCCACTCCTGTCGGGTGG-3' (SEQ ID NO:512)

Additionally, a synthetic oligonucleotide hybridization probe was constructed from the consensus DNA28735 sequence which had the following nucleotide sequence:

hybridization probe

5'-GGTGACACTTGCCAGTCAGATGTGGATGAATGCAGTGCTAGGAGGG-3' (SEQ ID NO:513)

In order to screen several libraries for a source of a full-length clone, DNA from the libraries was screened by PCR amplification with the PCR primer pair identified above. A positive library was then used to isolate clones encoding the PRO213-1, PRO1330 and/or PRO1449 gene using the probe oligonucleotide and one of the PCR primers. RNA for construction of the cDNA libraries was isolated from human fetal lung tissue.

DNA sequencing of the clones isolated as described above gave the full-length DNA sequence encoding PRO213-1, PRO1330 and/or PRO1449 [DNA30943-1-1163-1 (SEQ ID NO:505), DNA64907-1163-1 (SEQ ID NO:507) and DNA64908-1163-1 (SEQ ID NO:509), respectively].

The entire nucleotide sequences corresponding to DNA30943-1-1163-1 (SEQ ID NO:505), DNA64907-1163-1 (SEQ ID NO:507) and DNA64908-1163-1 (SEQ ID NO:509), respectively. DNA30943-1163, DNA64907-1163-1 and DNA64908-1163-1 contain a single open reading frame with an apparent translational initiation site at nucleotide positions 336-338, 488-490 and 326-328, respectively, and ending at the stop codon at nucleotide positions 1221-1223, 1307-1309 and 1145-1147, respectively (Figures 212, 214 and 216). The predicted polypeptide precursor is 295, 273 and 273 amino acids long, respectively (Figures 213, 215 and 217). DNA30943-1-1163-1, DNA64907-1163-1 and DNA64908-1163-1 have been deposited with ATCC and are assigned ATCC deposit no. 209791, 203242 and 203243, respectively.

Analysis of the amino acid sequence of the full-length PRO213-1 polypeptide suggests that a portion of it possess significant homology to the human growth arrest-specific gene 6 protein. More specifically, an analysis of the Dayhoff database (version 35.45 SwissProt 35) evidenced significant homology between the PRO213 amino acid sequence and the following Dayhoff sequences, HSMHC3W5A\_6 and B48089.

Additional analysis of the amino acid sequence of the full-length PRO1330 and PRO1449 polypeptide indicates significant identity with notch4. More specifically, an analysis of the Dayhoff database (version 35.130 SwissProt 35) evidenced significant identity between PRO1330 and the following Dayhoff sequences, D86566\_1 and NEL\_HUMAN.

#### EXAMPLE 95: Isolation of cDNA Clones Encoding Human PRO298

A cDNA isolated in the amylase screen described in Example 2 above is herein designated DNA26832 (Figure 220; SEQ ID NO:516). The sequence of DNA26832 was then used to search expressed sequence tag (EST) databases. The EST databases included public EST databases (e.g., GenBank) and a proprietary EST database (LIFESEQ™, Incyte Pharmaceuticals, Palo Alto, CA). The search was performed using the computer program BLAST or BLAST2 (Altschul et al., *Methods in Enzymology* 266: 469-480 [1996]). Those comparisons resulting in a BLAST score of 70 (or in some cases 90) or greater that did not encode proteins were clustered and assembled into consensus DNA sequences with the program "phrap" (Phil Green, University of Washington, Seattle, Washington; <http://bozeman.mbt.washington.edu/phrap.docs/phrap.html>).

A consensus DNA sequence was assembled relative to other EST sequences using phrap. A consensus sequence was determined, which was then extended using repeated cycles of BLAST and phrap to extend the consensus sequence as far as possible using the sources of EST sequences discussed above. The extended assembly sequence was designated DNA35861. Based on the DNA35861 consensus sequence, oligonucleotides were synthesized: 1) to identify by PCR a cDNA library that contained the sequence of interest, and 2) for use as probes to isolate a clone of the full-length coding sequence of PRO298. Forward and reverse primers

generally range from 20 to 30 nucleotides and are often designed to give a PCR product of about 100-1000 bp in length. The probe sequence is typically 40-55 bp in length. In some cases, additional oligonucleotides are synthesized when the consensus sequence is greater than about 1-1.5kbp. In order to screen several libraries for a full-length clone, DNA from the libraries was screened by PCR amplification, as per Ausubel et al., Current Protocols in Molecular Biology, with the PCR primer pair. A positive library was used to isolate clones

encoding the gene of interest using the probe oligonucleotide and one of the primer pairs.

PCR primers (forward and reverse) and a hybridization probe were synthesized:

forward PCR primer 1 CAACGTGATTTCAAAGCTGGGCTC (SEQ ID NO:517)

forward PCR primer 2 GCCTCGTATCAAGAAATTCC (SEQ ID NO:518)

forward PCR primer 3 AGTGGAGTCGACCTCCC (SEQ ID NO:519)

reverse PCR primer 1 CTCACCTGAAATCTCTCATAGCCC (SEQ ID NO:520)

hybridization probe 1 CGCAAAACCCATTTTGGGAGCAGGAATCCAATCATGTCTGTATGTTGG (SEQ ID NO:521)

In order to screen several libraries for a source of a full-length clone, DNA from the libraries was screened by PCR amplification with the PCR primer pair identified above. A positive library was then used to isolate clones encoding the PRO298 gene using the probe oligonucleotide and one of the PCR primers.

RNA for construction of the cDNA libraries was isolated from human fetal lung tissue (LIB25). The cDNA libraries used to isolate the cDNA clones were constructed by standard methods using commercially available reagents such as those from Invitrogen, San Diego, CA. The cDNA was primed with oligo dT containing a NotI site, linked with blunt to SalI hemikinased adaptors, cleaved with NotI, sized appropriately by gel electrophoresis, and cloned in a defined orientation into a suitable cloning vector (such as pRKB or pRKD; pRK5B is a precursor of pRK5D that does not contain the SfiI site; see, Holmes et al., Science, 253:1278-1280 (1991)) in the unique XhoI and NotI sites.

DNA sequencing of the clones isolated as described above gave the full-length DNA sequence for PRO298 (herein designated UNQ261 [DNA39975-1210]) (SEQ ID NO:514), and the derived protein sequence for PRO298 (SEQ ID NO:515).

The entire nucleotide sequence of UNQ261 (DNA39975-1210) is shown in Figure 218 (SEQ ID NO:514). Clone DNA39975-1210 contains a single open reading frame with an apparent translational initiation site at nucleotide positions 375-377. The predicted polypeptide precursor is 364 amino acids long. The protein contains four putative transmembrane domains between amino acid positions 36-55 (type II TM), 65-84, 188-208, and 229-245, respectively. A putative N-linked glycosylation site starts at amino acid position 253. In addition, the following features have been identified in the protein sequence: cAMP- and cGMP-dependent protein kinase phosphorylation site, starting at position 8; N-myristoylation sites starting at position 173 and 262, respectively; and a ZP domain between amino acid positions 45-60. Clone DNA39975-1210 has been deposited with ATCC (April 21, 1998) and is assigned ATCC deposit no.209783.

#### EXAMPLE 96: Isolation of cDNA Clones Encoding Human PRO337

A cDNA sequence identified in the amylase screen described in Example 2 above is herein designated DNA42301 (Figure 223, SEQ ID NO:524). The DNA42301 sequence was then compared to other EST

sequences using phrap as described in Example 1 above and a consensus sequence designated herein as DNA28761 was identified. Based on this consensus sequence, oligonucleotides were synthesized: 1) to identify by PCR a cDNA library that contained the sequence of interest, and 2) for use as probes to isolate a clone of the full-length coding sequence. In order to screen several libraries for a source of a full-length clone, DNA from the libraries was screened by PCR amplification with the PCR primer pair identified above. A positive library was then used to isolate clones encoding the PRO337 gene using the probe oligonucleotide and one of the PCR primers. RNA for construction of the cDNA libraries was isolated from human fetal brain.

A cDNA clone was sequenced in its entirety. The full length nucleotide sequence of DNA43316-1237 is shown in Figure 221 (SEQ ID NO:522). Clone DNA43316-1237 contains a single open reading frame with an apparent translational initiation site at nucleotide positions 134-136 (Figure 221; SEQ ID NO:522). The predicted polypeptide precursor is 344 amino acids long. Clone DNA43316-1237 has been deposited with ATCC and is assigned ATCC deposit no. 209487

Based on a BLAST-2 and FastA sequence alignment analysis of the full-length sequence, PRO337 shows amino acid sequence identity to rat neurotrimin (97%).

#### EXAMPLE 97: Isolation of cDNA Clones Encoding Human PRO403

##### Introduction:

Human thrombopoietin (THPO) is a glycosylated hormone of 352 amino acids consisting of two domains. The N-terminal domain, sharing 50% similarity to erythropoietin, is responsible for the biological activity. The C-terminal region is required for secretion. The gene for thrombopoietin (THPO) maps to human chromosome 3q27-q28 where the six exons of this gene span 7 kilobase base pairs of genomic DNA (Chang et al., Genomics 26: 636-7 (1995); Foster et al., Proc. Natl. Acad. Sci. USA 91: 13023-7 (1994); Gurney et al., Blood 85: 981-988 (1995). In order to determine whether there were any genes encoding THPO homologues located in close proximity to THPO, genomic DNA fragments from this region were identified and sequenced. Three P1 clones and one PAC clones (Genome Systems Inc., St. Louis, MO; cat. Nos. P1-2535 and PAC-6539) encompassing the THPO locus were isolated and a 140 kb region was sequenced using the ordered shotgun strategy (Chen et al., Genomics 17: 651-656 (1993)), coupled with a PCR-based gap filling approach. Analysis reveals that the region is gene-rich with four additional genes located very close to THPO: tumor necrosis factor-receptor type 1 associated protein 2 (TRAP2) and elongation initiation factor gamma (eIF4 $\gamma$ ), chloride channel 2 (CLCN2) and RNA polymerase II subunit hRPB17. While no THPO homolog was found in the region, four novel genes have been predicted by computer-assisted gene detection (GRAIL)(Xu et al., Gen. Engin. 16: 241-253 (1994), the presence of CpG islands (Cross, S. and Bird, A., Curr. Opin. Genet. & Devel. 5: 109-314 (1995), and homology to known genes (as detected by WU-BLAST2.0)(Altschul and Gish, Methods Enzymol. 266: 460-480 (1996) (<http://blast.wustl.edu/blast/README.html>).

##### Procedures:

##### P1 and PAC clones:

The initial human P1 clone was isolated from a genomic P1 library (Genome Systems Inc., St. Louis, MO; cat. no.: P1-2535) screened with PCR primers designed from the THPO genomic sequence (A.L. Gurney, et al., Blood 85: 981-88 (1995). PCR primers were designed from the end sequences derived from this P1 clone

were then used to screen P1 and PAC libraries (Genome Systems, Cat. Nos.: P1-2535 & PAC-6539) to identify overlapping clones (PAC1, p1.t, and P1.u). The 3'-end sequence from PAC.z was used to define the primers used for the screening of a human BAC library (Genome Systems Inc., St. Louis, MO; Cat. No.: BDTW-4533A).

#### Ordered Shotgun Strategy:

The Ordered Shotgun Strategy (OSS) (Chen et al., Genomics 17: 651-656 (1993)) involves the mapping and sequencing of large genomic DNA clones with a hierarchical approach. The P1 or PAC clone was sonicated and the fragments subcloned into lambda vector ( $\lambda$ Bluestar) (Novagen, Inc., Madison, WI; cat. no. 69242-3). The lambda subclone inserts were isolated by long-range PCR (Barnes, W. Proc. Natl. Acad. Sci. USA 91: 2216-2220 (1994) and the ends sequenced. The lambda-end sequences were overlapped to create a partial map of the original clone. Those lambda clones with overlapping end-sequences were identified, the inserts subcloned into a plasmid vector (pUC18 or pUC19, Hoefer Pharmacia Biotech, Inc., San Francisco, CA, Cat. Nos. 27-4949-01 and 27-4951-01) and the ends of the plasmid subclones were sequenced and assembled to generate a contiguous sequence. This directed sequencing strategy minimizes the redundancy required while allowing one to scan for and concentrate on interesting regions.

In order to define better the THPO locus and to search for other genes related to the hematopoietin family, five genomic clones were isolated from this region by PCR screening of human P1 and PAC libraries (Genome System, Inc., Cat. Nos.: P1-2535 and PAC-6539).

The sizes of the genomic fragments are as follows: P1.t is 40 kb; P1.g is 70 kb; P1.u is 70 kb; PAC.z is 200 kb; and BAC.1 is 80 kb. Approximately 75% (140 kb) of the 190 kb genomic DNA region was sequenced by the Ordered Shotgun Strategy (OSS) (Chen et al., Genomics 17: 651-56 (1993), and assembled into contigs using AutoAssembler<sup>TM</sup> (Applied Biosystems, Perkin Elmer, Foster City, CA, cat. no. 903227). The preliminary order of these contigs was determined by manual analysis. There were 47 contigs the 140 kb region. A PCR-based approach to ordering the contigs and filling in the gaps was employed. The following summarizes the number and sizes of the gaps. The 50 kb of sequence unique to BAC.1 was sequenced by a total shotgun approach with a ten-fold redundancy.

<u>Size of gap</u>	<u>number</u>
<50 bp	13
50-150 bp	7
150-300 bp	7
300-1000 bp	10
1000-5000 bp	7
> 5000 bp	2 ((15,000 bp)

#### DNA sequencing:

ABIDYE-primer<sup>TM</sup> chemistry (PE Applied Biosystems, Foster City, CA; Cat. No.: 402112) was used to end-sequence the lambda and plasmid subclones. ABIDYE-terminator<sup>TM</sup> chemistry (PE Applied Biosystems, Foster City, CA, Cat. No: 403044) was used to sequence the PCR products with their respective PCR primers. The sequences were collected with an ABI377 instrument. For PCR products larger than 1kb, walking primers

were used. The sequences of contigs generated by the OSS strategy in AutoAssemblerTM (PE Applied Biosystems, Foster City, CA; Cat. No: 903227) and the gap-filling sequencing trace files were imported into SequencherTM (Gene Codes Corp., Ann Arbor, MI) for overlapping and editing. The sequences generated by the total shotgun strategy were assembled using Phred and Phrap and edited using Consed (<http://chimera.biotech.washington.edu/uwgc/projects.htm>) and GFP (Genome Reconstruction Manager for Phrap), version 1.2 (<http://stork.cellb.bcm.tmc.edu/gfp/>).

#### PCR-Based gap filling Strategy:

Primers were designed based on the 5'- and 3'-end sequenced of each contig, avoiding repetitive and low quality sequence regions. All primers were designed to be 19-24-mers with 50-70% G/C content. Oligos were synthesized and gel-purified by standard methods.

Since the orientation and order of the contigs were unknown, permutations of the primers were used in the amplification reactions. Two PCR kits were used: first, XL PCR kit (Perkin Elmer, Norwalk, CT; Cat. No.: N8080205), with extension times of approximately 10 minutes; and second, the Taq polymerase PCR kit (Qiagen Inc., Valencia, CA; Cat. No.: 201223) was used under high stringency conditions if smeared or multiple products were observed with the XL PCR kit. The main PCR product from each successful reaction was extracted from a 0.9% low melting agarose gel and purified with the Geneclean DNA Purification kit prior to sequencing.

#### Analysis:

The identification and characterization of coding regions was carried out as follows: First, repetitive sequences were masked using RepeatMasker (A.F.A. Smit & P. Green, [http://ftp.genome.washington.edu/RM/RM\\_details.html](http://ftp.genome.washington.edu/RM/RM_details.html)) which screens DNA sequences in FastA format against a library of repetitive elements and returns a masked query sequence. Repeats not masked were identified by comparing the sequence to the GenBank database using WUBLAST2.0 [Altschul, S & Gish, W., Methods Enzymol. 266: 460-480 (1996); <http://blast.wustl.edu/blast/README.html>] and were masked manually.

Next, known genes were revealed by comparing the genomic regions against Genentech's protein database using the WUBLAST2.0 algorithm and then annotated by aligning the genomic and cDNA sequences for each gene, respectively, using a Needleman-Wunch (Needleman and Wunch, J. Mol. Biol. 48: 443-453 (1970) algorithm to find regions of local identity between sequences. The strategy results in detection of all exons of the five known genes in the region, THPO, TRAP2, eIF4g, CLCN2 and hRPB17 (see below).

#### Known genes

eukaryotic translation initiation factor 4 gamma  
thrombopoietin  
chloride channel 2  
TNF receptor associated protein 2  
RNA polymerase II subunit hRPB17

#### Map position

3q27-qter  
3q26-q27  
3q26-qter  
not previously mapped  
not previously mapped

Finally, novel transcription units were predicted using a number of approaches. CpG islands (S. Cross & Bird, A., Curr. Opin. Genet. Dev. 5: 109-314 (1995) islands were used to define promoter regions and were

identified as clusters of sites cleaved by enzymes recognizing GC-rich, 6 or 8-mer palindromic sequences (NotI, NarI, BssHII, XhoI. CpG islands are usually associated with promoter regions of genes. WUBLAST2.0 analysis of short genomic regions (10-20 kb) versus GenBank revealed matches to ESTs. The individual EST sequences (or where possible, their sequence chromatogram files) were retrieved and assembled with Sequencer to provide a theoretical cDNA sequence (DNA36443). GRAIL2 (ApoCom Inc., Knoxville, TN, command line version for the DEC alpha) was used to predict a novel exon. The five known genes in the region served as internal controls for the success of the GRAIL algorithm.

#### Isolation:

A partial endothelin converting enzyme-2 (ECE-2) cDNA clone was isolated by first splicing in silico the ECE-2 exons predicted in the genomic sequence to generate a putative sequence (DNA36443). An oligonucleotide probe: GAAGCAGTGCAGCCAGCAGTAGAGAGGCACCTGCTAAGA) (SEQ ID NO:530) was designed and used to screen a human fetal small intestine library (LIB110) and internal PCR primers (36443f1) (ECE2.f:ACGCAGCTGGAGCTGGTCTTAGCA) (SEQ ID NO:531) and (36443r1) (ECE2.r) (GGTACTGGACCCCTAGGGCCACAA) (SEQ ID NO:532) were used to confirm clones hybridizing to the probe prior to sequencing. One positive clone was obtained, however this cDNA (DNA49830) represented a partially spliced transcript containing appropriately spliced exons 1 through 6, followed by intron 6 sequence. The oligo dT primer annealed to a polyA-stretch within an Alu element present in intron 6. An additional ECE-2 cDNA fragment (DNA49831) was obtained by PCR from a human fetal kidney library (LIB227) with primers designed from the presumed cDNA sequence [36443f3: CCTCCCAGCCGAGACAGTGG (SEQ ID NO:533) and 36443r2: GGTCTATAAGGGCCAAGACC (SEQ ID NO:534)]. This PCR product extended from exon 13 into the 3' untranslated region in exon 18.

A full length endothelin converting enzyme 2 (ECE-2) cDNA clone (DNA55800-1263) was isolated from an oligo-dT-primed human fetal brain library. RNA from human fetal brain tissue (20 weeks gestation, #283005)(SRC175) was isolated by guanidine thiocyanate and 5 µg used to generate double stranded cDNA which was cloned into the vector pRK5E. The 3' -primer (pGACTAGTTCTAGATCGCGAGCGGCCGCCCTTTTTTTTTTTTTT) (SEQ ID NO:535) and the 5 -linker (pCGGACGCGTGGGTCGA) (SEQ ID NO:536) were designed to introduce XhoI and NotI restriction sites. The library was screened with PCR primers [36443pcrf1: CGGCCGTGATGGCTGGTGACG (SEQ ID NO:537) and 36443r3: GGCAGACTCCTTCCTATGGG (SEQ ID NO:538)] designed from the partial human ECE-2 cDNA sequences (DNA49830 and DNA49831). PCR products were cloned into the vector pCR2.1-TOPO (Invitrogen Corp., Carlsbad, CA, Cat. No. K4500-01) and sequenced with DYE-terminator chemistry as described above.

#### EXAMPLE 98: Northern Blot and in situ RNA Hybridization Analysis for PRO403

Expression of PRO403 mRNA in human tissues was examined by Northern blot analysis. Human polyA+ RNA blots derived from human fetal and adult tissues (Clontech, Palo Alto, CA; Cat. Nos. 7760-1, 7756-1 and 7755-1) were hybridized to a [32P-α]dATP-labelled cDNA fragments from probe based on the full length PRO403 cDNA. Blots were incubated with the probes in hybridization buffer (5X SSPE; 2X Denhardt's solution; 100 mg/mL denatured sheared salmon sperm DNA; 50% formamide; 2% SDS) for 18 hours at 42°C,

washed to high stringency (0.1XSSC, 0.1% SDS, 50°C) and autoradiographed. The blots were developed after overnight exposure by phosphorimager analysis (Fuji).

PRO403 mRNA transcripts were detected. Analysis of the expression pattern showed the strongest signal of the expected 3.3 kb transcript in adult brain (highest in the cerebellum, putamen, medulla, and temporal lobe, and lower in the cerebral cortex, occipital lobe and frontal lobe), spinal cord, lung and pancreas and higher levels of a 4.5 kb transcript in fetal brain and kidney.

#### EXAMPLE 99: Use of PRO Polypeptide-Encoding Nucleic Acid as Hybridization Probes

The following method describes use of a nucleotide sequence encoding a PRO polypeptide as a hybridization probe.

DNA comprising the coding sequence of a PRO polypeptide of interest as disclosed herein may be employed as a probe or used as a basis from which to prepare probes to screen for homologous DNAs (such as those encoding naturally-occurring variants of the PRO polypeptide) in human tissue cDNA libraries or human tissue genomic libraries.

Hybridization and washing of filters containing either library DNAs is performed under the following high stringency conditions. Hybridization of radiolabeled PRO polypeptide-encoding nucleic acid-derived probe to the filters is performed in a solution of 50% formamide, 5x SSC, 0.1% SDS, 0.1% sodium pyrophosphate, 50 mM sodium phosphate, pH 6.8, 2x Denhardt's solution, and 10% dextran sulfate at 42°C for 20 hours. Washing of the filters is performed in an aqueous solution of 0.1x SSC and 0.1% SDS at 42°C.

DNAs having a desired sequence identity with the DNA encoding full-length native sequence PRO polypeptide can then be identified using standard techniques known in the art.

#### EXAMPLE 100: Expression of PRO Polypeptides in *E. coli*

This example illustrates preparation of an unglycosylated form of a desired PRO polypeptide by recombinant expression in *E. coli*.

The DNA sequence encoding the desired PRO polypeptide is initially amplified using selected PCR primers. The primers should contain restriction enzyme sites which correspond to the restriction enzyme sites on the selected expression vector. A variety of expression vectors may be employed. An example of a suitable vector is pBR322 (derived from *E. coli*; see Bolivar et al., *Gene*, 2:95 (1977)) which contains genes for ampicillin and tetracycline resistance. The vector is digested with restriction enzyme and dephosphorylated. The PCR amplified sequences are then ligated into the vector. The vector will preferably include sequences which encode for an antibiotic resistance gene, a trp promoter, a polyhis leader (including the first six STII codons, polyhis sequence, and enterokinase cleavage site), the specific PRO polypeptide coding region, lambda transcriptional terminator, and an argU gene.

The ligation mixture is then used to transform a selected *E. coli* strain using the methods described in Sambrook et al., *supra*. Transformants are identified by their ability to grow on LB plates and antibiotic resistant colonies are then selected. Plasmid DNA can be isolated and confirmed by restriction analysis and DNA sequencing.

Selected clones can be grown overnight in liquid culture medium such as LB broth supplemented with antibiotics. The overnight culture may subsequently be used to inoculate a larger scale culture. The cells are then grown to a desired optical density, during which the expression promoter is turned on.

After culturing the cells for several more hours, the cells can be harvested by centrifugation. The cell pellet obtained by the centrifugation can be solubilized using various agents known in the art, and the solubilized PRO polypeptide can then be purified using a metal chelating column under conditions that allow tight binding of the protein.

PRO181, PRO195, PRO200, PRO237, PRO273, PRO540, PRO322, PRO1017, PRO938, PRO162, PRO1114, PRO827 and PRO1008 were expressed in *E. coli* in a poly-His tagged form, using the following procedure. The DNA encoding the PRO polypeptide was initially amplified using selected PCR primers. The primers contained restriction enzyme sites which correspond to the restriction enzyme sites on the selected expression vector, and other useful sequences providing for efficient and reliable translation initiation, rapid purification on a metal chelation column, and proteolytic removal with enterokinase. The PCR-amplified, poly-His tagged sequences were then ligated into an expression vector, which was used to transform an *E. coli* host based on strain 52 (W3110 fuhA(tonA) lon galE rpoHts(htpRts) clpP(lacIq)). Transformants were first grown in LB containing 50 mg/ml carbenicillin at 30°C with shaking until an O.D.600 of 3-5 was reached. Cultures were then diluted 50-100 fold into CRAP media (prepared by mixing 3.57 g (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 0.71 g sodium citrate-2H<sub>2</sub>O, 1.07 g KCl, 5.36 g Difco yeast extract, 5.36 g Sheffield hycase SF in 500 mL water, as well as 110 mM MPOS, pH 7.3, 0.55% (w/v) glucose and 7 mM MgSO<sub>4</sub>) and grown for approximately 20-30 hours at 30°C with shaking. Samples were removed to verify expression by SDS-PAGE analysis, and the bulk culture is centrifuged to pellet the cells. Cell pellets were frozen until purification and refolding.

*E. coli* paste from 0.5 to 1 L fermentations (6-10 g pellets) was resuspended in 10 volumes (w/v) in 7 M guanidine, 20 mM Tris, pH 8 buffer. Solid sodium sulfite and sodium tetrathionate is added to make final concentrations of 0.1M and 0.02 M, respectively, and the solution was stirred overnight at 4°C. This step results in a denatured protein with all cysteine residues blocked by sulfitolization. The solution was centrifuged at 40,000 rpm in a Beckman Ultracentrifuge for 30 min. The supernatant was diluted with 3-5 volumes of metal chelate column buffer (6 M guanidine, 20 mM Tris, pH 7.4) and filtered through 0.22 micron filters to clarify. Depending the clarified extract was loaded onto a 5 ml Qiagen Ni-NTA metal chelate column equilibrated in the metal chelate column buffer. The column was washed with additional buffer containing 50 mM imidazole (Calbiochem, Utrol grade), pH 7.4. The protein was eluted with buffer containing 250 mM imidazole. Fractions containing the desired protein were pooled and stored at 4°C. Protein concentration was estimated by its absorbance at 280 nm using the calculated extinction coefficient based on its amino acid sequence.

The proteins were refolded by diluting sample slowly into freshly prepared refolding buffer consisting of: 20 mM Tris, pH 8.6, 0.3 M NaCl, 2.5 M urea, 5 mM cysteine, 20 mM glycine and 1 mM EDTA. Refolding volumes were chosen so that the final protein concentration was between 50 to 100 micrograms/ml. The refolding solution was stirred gently at 4°C for 12-36 hours. The refolding reaction was quenched by the addition of TFA to a final concentration of 0.4% (pH of approximately 3). Before further purification of the protein, the solution was filtered through a 0.22 micron filter and acetonitrile was added to 2-10% final concentration. The refolded protein was chromatographed on a Poros R1/H reversed phase column using a

mobile buffer of 0.1% TFA with elution with a gradient of acetonitrile from 10 to 80%. Aliquots of fractions with A280 absorbance were analyzed on SDS polyacrylamide gels and fractions containing homogeneous refolded protein were pooled. Generally, the properly refolded species of most proteins are eluted at the lowest concentrations of acetonitrile since those species are the most compact with their hydrophobic interiors shielded from interaction with the reversed phase resin. Aggregated species are usually eluted at higher acetonitrile concentrations. In addition to resolving misfolded forms of proteins from the desired form, the reversed phase step also removes endotoxin from the samples.

Fractions containing the desired folded PRO proteins were pooled and the acetonitrile removed using a gentle stream of nitrogen directed at the solution. Proteins were formulated into 20 mM Hepes, pH 6.8 with 0.14 M sodium chloride and 4% mannitol by dialysis or by gel filtration using G25 Superfine (Pharmacia) resins equilibrated in the formulation buffer and sterile filtered.

Many of the PRO polypeptides described herein were successfully expressed as described above.

#### EXAMPLE 101: Expression of PRO Polypeptides in Mammalian Cells

This example illustrates preparation of a glycosylated form of a desired PRO polypeptide by recombinant expression in mammalian cells.

The vector, pRK5 (see EP 307,247, published March 15, 1989), is employed as the expression vector. Optionally, the PRO polypeptide-encoding DNA is ligated into pRK5 with selected restriction enzymes to allow insertion of the PRO polypeptide DNA using ligation methods such as described in Sambrook et al., *supra*. The resulting vector is called pRK5-PRO polypeptide.

In one embodiment, the selected host cells may be 293 cells. Human 293 cells (ATCC CCL 1573) are grown to confluence in tissue culture plates in medium such as DMEM supplemented with fetal calf serum and optionally, nutrient components and/or antibiotics. About 10  $\mu$ g pRK5-PRO polypeptide DNA is mixed with about 1  $\mu$ g DNA encoding the VA RNA gene [Thimmappaya et al., *Cell*, 31:543 (1982)] and dissolved in 500  $\mu$ l of 1 mM Tris-HCl, 0.1 mM EDTA, 0.227 M  $\text{CaCl}_2$ . To this mixture is added, dropwise, 500  $\mu$ l of 50 mM HEPES (pH 7.35), 280 mM NaCl, 1.5 mM  $\text{NaPO}_4$ , and a precipitate is allowed to form for 10 minutes at 25°C. The precipitate is suspended and added to the 293 cells and allowed to settle for about four hours at 37°C. The culture medium is aspirated off and 2 ml of 20% glycerol in PBS is added for 30 seconds. The 293 cells are then washed with serum free medium, fresh medium is added and the cells are incubated for about 5 days.

Approximately 24 hours after the transfections, the culture medium is removed and replaced with culture medium (alone) or culture medium containing 200  $\mu$ Cl/ml  $^{35}\text{S}$ -cysteine and 200  $\mu$ Cl/ml  $^{35}\text{S}$ -methionine. After a 12 hour incubation, the conditioned medium is collected, concentrated on a spin filter, and loaded onto a 15% SDS gel. The processed gel may be dried and exposed to film for a selected period of time to reveal the presence of PRO polypeptide. The cultures containing transfected cells may undergo further incubation (in serum free medium) and the medium is tested in selected bioassays.

In an alternative technique, PRO polypeptide may be introduced into 293 cells transiently using the dextran sulfate method described by Somparyrac et al., *Proc. Natl. Acad. Sci.*, 12:7575 (1981). 293 cells are grown to maximal density in a spinner flask and 700  $\mu$ g pRK5-PRO polypeptide DNA is added. The cells are first concentrated from the spinner flask by centrifugation and washed with PBS. The DNA-dextran precipitate

In order to screen several libraries for a source of a full-length clone, DNA from the libraries was screened by PCR amplification with the PCR primer pair identified above. A positive library was then used to isolate clones encoding the PRO213-1, PRO1330 and/or PRO1449 gene using the probe oligonucleotide and one of the PCR primers. RNA for construction of the cDNA libraries was isolated from human fetal lung tissue.

DNA sequencing of the clones isolated as described above gave the full-length DNA sequence encoding PRO213-1, PRO1330 and/or PRO1449 [DNA30943-1-1163-1 (SEQ ID NO:505), DNA64907-1163-1 (SEQ ID NO:507) and DNA64908-1163-1 (SEQ ID NO:509), respectively].

The entire nucleotide sequences corresponding to DNA30943-1-1163-1 (SEQ ID NO:505), DNA64907-1163-1 (SEQ ID NO:507) and DNA64908-1163-1 (SEQ ID NO:509), respectively. DNA30943-1163, DNA64907-1163-1 and DNA64908-1163-1 contain a single open reading frame with an apparent translational initiation site at nucleotide positions 336-338, 488-490 and 326-328, respectively, and ending at the stop codon at nucleotide positions 1221-1223, 1307-1309 and 1145-1147, respectively (Figures 212, 214 and 216). The predicted polypeptide precursor is 295, 273 and 273 amino acids long, respectively (Figures 213, 215 and 217). DNA30943-1-1163-1, DNA64907-1163-1 and DNA64908-1163-1 have been deposited with ATCC and are assigned ATCC deposit no. 209791, 203242 and 203243, respectively.

Analysis of the amino acid sequence of the full-length PRO213-1 polypeptide suggests that a portion of it possess significant homology to the human growth arrest-specific gene 6 protein. More specifically, an analysis of the Dayhoff database (version 35.45 SwissProt 35) evidenced significant homology between the PRO213 amino acid sequence and the following Dayhoff sequences, HSMHC3W5A\_6 and B48089.

Additional analysis of the amino acid sequence of the full-length PRO1330 and PRO1449 polypeptide indicates significant identity with notch4. More specifically, an analysis of the Dayhoff database (version 35.130 SwissProt 35) evidenced significant identity between PRO1330 and the following Dayhoff sequences, D86566\_1 and NEL\_HUMAN.

#### EXAMPLE 95: Isolation of cDNA Clones Encoding Human PRO298

A cDNA isolated in the amylase screen described in Example 2 above is herein designated DNA26832 (Figure 220; SEQ ID NO:516). The sequence of DNA26832 was then used to search expressed sequence tag (EST) databases. The EST databases included public EST databases (e.g., GenBank) and a proprietary EST database (LIFESEQ™, Incyte Pharmaceuticals, Palo Alto, CA). The search was performed using the computer program BLAST or BLAST2 (Altschul et al., *Methods in Enzymology* 266: 469-480 [1996]). Those comparisons resulting in a BLAST score of 70 (or in some cases 90) or greater that did not encode proteins were clustered and assembled into consensus DNA sequences with the program "phrap" (Phil Green, University of Washington, Seattle, Washington; <http://bozeman.mbt.washington.edu/phrap.docs/phrap.html>).

A consensus DNA sequence was assembled relative to other EST sequences using phrap. A consensus sequence was determined, which was then extended using repeated cycles of BLAST and phrap to extend the consensus sequence as far as possible using the sources of EST sequences discussed above. The extended assembly sequence was designated DNA35861. Based on the DNA35861 consensus sequence, oligonucleotides were synthesized: 1) to identify by PCR a cDNA library that contained the sequence of interest, and 2) for use as probes to isolate a clone of the full-length coding sequence of PRO298. Forward and reverse primers

generally range from 20 to 30 nucleotides and are often designed to give a PCR product of about 100-1000 bp in length. The probe sequence is typically 40-55 bp in length. In some cases, additional oligonucleotides are synthesized when the consensus sequence is greater than about 1-1.5kbp. In order to screen several libraries for a full-length clone, DNA from the libraries was screened by PCR amplification, as per Ausubel et al., Current Protocols in Molecular Biology, with the PCR primer pair. A positive library was used to isolate clones

encoding the gene of interest using the probe oligonucleotide and one of the primer pairs.

PCR primers (forward and reverse) and a hybridization probe were synthesized:

forward PCR primer 1 CAACGTGATTCAAAGCTGGGCTC (SEQ ID NO:517)

forward PCR primer 2 GCCTCGTATCAAGAAATTCC (SEQ ID NO:518)

forward PCR primer 3 AGTGGAAGTCGACCTCCC (SEQ ID NO:519)

reverse PCR primer 1 CTCACCTGAAATCTCTCATAGCCC (SEQ ID NO:520)

hybridization probe 1 CGCAAAACCCATTTTGGGAGCAGGAATTCCAATCATGTCTGTGATGGTGG (SEQ ID NO:521)

In order to screen several libraries for a source of a full-length clone, DNA from the libraries was screened by PCR amplification with the PCR primer pair identified above. A positive library was then used to isolate clones encoding the PRO298 gene using the probe oligonucleotide and one of the PCR primers.

RNA for construction of the cDNA libraries was isolated from human fetal lung tissue (LIB25). The cDNA libraries used to isolate the cDNA clones were constructed by standard methods using commercially available reagents such as those from Invitrogen, San Diego, CA. The cDNA was primed with oligo dT containing a NotI site, linked with blunt to SalI hemikinased adaptors, cleaved with NotI, sized appropriately by gel electrophoresis, and cloned in a defined orientation into a suitable cloning vector (such as pRKB or pRKD; pRK5B is a precursor of pRK5D that does not contain the SfiI site; see, Holmes et al., Science, 253:1278-1280 (1991)) in the unique XhoI and NotI sites.

DNA sequencing of the clones isolated as described above gave the full-length DNA sequence for PRO298 (herein designated UNQ261 [DNA39975-1210]) (SEQ ID NO:514), and the derived protein sequence for PRO298 (SEQ ID NO:515).

The entire nucleotide sequence of UNQ261 (DNA39975-1210) is shown in Figure 218 (SEQ ID NO:514). Clone DNA39975-1210 contains a single open reading frame with an apparent translational initiation site at nucleotide positions 375-377. The predicted polypeptide precursor is 364 amino acids long. The protein contains four putative transmembrane domains between amino acid positions 36-55 (type II TM), 65-84, 188-208, and 229-245, respectively. A putative N-linked glycosylation site starts at amino acid position 253. In addition, the following features have been identified in the protein sequence: cAMP- and cGMP-dependent protein kinase phosphorylation site, starting at position 8; N-myristoylation sites starting at position 173 and 262, respectively; and a ZP domain between amino acid positions 45-60. Clone DNA39975-1210 has been deposited with ATCC (April 21, 1998) and is assigned ATCC deposit no.209783.

#### EXAMPLE 96: Isolation of cDNA Clones Encoding Human PRO337

A cDNA sequence identified in the amylase screen described in Example 2 above is herein designated DNA42301 (Figure 223, SEQ ID NO:524). The DNA42301 sequence was then compared to other EST

sequences using phrap as described in Example 1 above and a consensus sequence designated herein as DNA28761 was identified. Based on this consensus sequence, oligonucleotides were synthesized: 1) to identify by PCR a cDNA library that contained the sequence of interest, and 2) for use as probes to isolate a clone of the full-length coding sequence. In order to screen several libraries for a source of a full-length clone, DNA from the libraries was screened by PCR amplification with the PCR primer pair identified above. A positive library was then used to isolate clones encoding the PRO337 gene using the probe oligonucleotide and one of the PCR primers. RNA for construction of the cDNA libraries was isolated from human fetal brain.

A cDNA clone was sequenced in its entirety. The full length nucleotide sequence of DNA43316-1237 is shown in Figure 221 (SEQ ID NO:522). Clone DNA43316-1237 contains a single open reading frame with an apparent translational initiation site at nucleotide positions 134-136 (Figure 221; SEQ ID NO:522). The predicted polypeptide precursor is 344 amino acids long. Clone DNA43316-1237 has been deposited with ATCC and is assigned ATCC deposit no. 209487

Based on a BLAST-2 and FastA sequence alignment analysis of the full-length sequence, PRO337 shows amino acid sequence identity to rat neurotrimin (97%).

#### EXAMPLE 97: Isolation of cDNA Clones Encoding Human PRO403

##### Introduction:

Human thrombopoietin (THPO) is a glycosylated hormone of 352 amino acids consisting of two domains. The N-terminal domain, sharing 50% similarity to erythropoietin, is responsible for the biological activity. The C-terminal region is required for secretion. The gene for thrombopoietin (THPO) maps to human chromosome 3q27-q28 where the six exons of this gene span 7 kilobase base pairs of genomic DNA (Chang et al., Genomics 26: 636-7 (1995); Foster et al., Proc. Natl. Acad. Sci. USA 91: 13023-7 (1994); Gurney et al., Blood 85: 981-988 (1995). In order to determine whether there were any genes encoding THPO homologues located in close proximity to THPO, genomic DNA fragments from this region were identified and sequenced. Three P1 clones and one PAC clones (Genome Systems Inc., St. Louis, MO; cat. Nos. P1-2535 and PAC-6539) encompassing the THPO locus were isolated and a 140 kb region was sequenced using the ordered shotgun strategy (Chen et al., Genomics 17: 651-656 (1993)), coupled with a PCR-based gap filling approach. Analysis reveals that the region is gene-rich with four additional genes located very close to THPO: tumor necrosis factor-receptor type 1 associated protein 2 (TRAP2) and elongation initiation factor gamma (eIF40), chloride channel 2 (CLCN2) and RNA polymerase II subunit hRPB17. While no THPO homolog was found in the region, four novel genes have been predicted by computer-assisted gene detection (GRAIL)(Xu et al., Gen. Engin. 16: 241-253 (1994), the presence of CpG islands (Cross, S. and Bird, A., Curr. Opin. Genet. & Devel. 5: 109-314 (1995), and homology to known genes (as detected by WU-BLAST2.0)(Altschul and Gish, Methods Enzymol. 266: 460-480 (1996) (<http://blast.wustl.edu/blast/README.html>).

##### Procedures:

##### P1 and PAC clones:

The initial human P1 clone was isolated from a genomic P1 library (Genome Systems Inc., St. Louis, MO; cat. no.: P1-2535) screened with PCR primers designed from the THPO genomic sequence (A.L. Gurney, et al., Blood 85: 981-88 (1995). PCR primers were designed from the end sequences derived from this P1 clone

were then used to screen P1 and PAC libraries (Genome Systems, Cat. Nos.: P1-2535 & PAC-6539) to identify overlapping clones (PAC1, p1.t, and P1.u). The 3'-end sequence from PAC.z was used to define the primers used for the screening of a human BAC library (Genome Systems Inc., St. Louis, MO; Cat. No.: BDTW-4533A).

#### Ordered Shotgun Strategy:

The Ordered Shotgun Strategy (OSS) (Chen et al., Genomics 17: 651-656 (1993)) involves the mapping and sequencing of large genomic DNA clones with a hierarchical approach. The P1 or PAC clone was sonicated and the fragments subcloned into lambda vector ( $\lambda$ BlueStar) (Novagen, Inc., Madison, WI; cat. no. 69242-3). The lambda subclone inserts were isolated by long-range PCR (Barnes, W. Proc. Natl. Acad. Sci. USA 91: 2216-2220 (1994) and the ends sequenced. The lambda-end sequences were overlapped to create a partial map of the original clone. Those lambda clones with overlapping end-sequences were identified, the inserts subcloned into a plasmid vector (pUC18 or pUC19, Hoefer Pharmacia Biotech, Inc., San Francisco, CA, Cat. Nos. 27-4949-01 and 27-4951-01) and the ends of the plasmid subclones were sequenced and assembled to generate a contiguous sequence. This directed sequencing strategy minimizes the redundancy required while allowing one to scan for and concentrate on interesting regions.

In order to define better the THPO locus and to search for other genes related to the hematopoietin family, five genomic clones were isolated from this region by PCR screening of human P1 and PAC libraries (Genome System, Inc., Cat. Nos.: P1-2535 and PAC-6539).

The sizes of the genomic fragments are as follows: P1.t is 40 kb; P1.g is 70 kb; P1.u is 70 kb; PAC.z is 200 kb; and BAC.1 is 80 kb. Approximately 75% (140 kb) of the 190 kb genomic DNA region was sequenced by the Ordered Shotgun Strategy (OSS) (Chen et al., Genomics 17: 651-56 (1993), and assembled into contigs using AutoAssembler<sup>TM</sup> (Applied Biosystems, Perkin Elmer, Foster City, CA, cat. no. 903227). The preliminary order of these contigs was determined by manual analysis. There were 47 contigs the 140 kb region. A PCR-based approach to ordering the contigs and filling in the gaps was employed. The following summarizes the number and sizes of the gaps. The 50 kb of sequence unique to BAC.1 was sequenced by a total shotgun approach with a ten-fold redundancy.

<u>Size of gap</u>	<u>number</u>
<50 bp	13
50-150 bp	7
150-300 bp	7
300-1000 bp	10
1000-5000 bp	7
> 5000 bp	2 ((15,000 bp)

#### DNA sequencing:

ABI DYE-primer<sup>TM</sup> chemistry (PE Applied Biosystems, Foster City, CA; Cat. No.: 402112) was used to end-sequence the lambda and plasmid subclones. ABI DYE-terminator<sup>TM</sup> chemistry (PE Applied Biosystems, Foster City, CA, Cat. No: 403044) was used to sequence the PCR products with their respective PCR primers. The sequences were collected with an ABI377 instrument. For PCR products larger than 1kb, walking primers

were used. The sequences of contigs generated by the OSS strategy in AutoAssemblerTM (PE Applied Biosystems, Foster City, CA; Cat. No: 903227) and the gap-filling sequencing trace files were imported into SequencerTM (Gene Codes Corp., Ann Arbor, MI) for overlapping and editing. The sequences generated by the total shotgun strategy were assembled using Phred and Phrap and edited using Consed (<http://chimera.biotech.washington.edu/uwgc/projects.htm>) and GFP (Genome Reconstruction Manager for Phrap), version 1.2 (<http://stork.cellb.bcm.tmc.edu/gfp/>).

#### PCR-Based gap filling Strategy:

Primers were designed based on the 5'- and 3'-end sequenced of each contig, avoiding repetitive and low quality sequence regions. All primers were designed to be 19-24-mers with 50-70% G/C content. Oligos were synthesized and gel-purified by standard methods.

Since the orientation and order of the contigs were unknown, permutations of the primers were used in the amplification reactions. Two PCR kits were used: first, XL PCR kit (Perkin Elmer, Norwalk, CT; Cat. No.: N8080205), with extension times of approximately 10 minutes; and second, the Taq polymerase PCR kit (Qiagen Inc., Valencia, CA; Cat. No.: 201223) was used under high stringency conditions if smeared or multiple products were observed with the XL PCR kit. The main PCR product from each successful reaction was extracted from a 0.9% low melting agarose gel and purified with the Geneclean DNA Purification kit prior to sequencing.

#### Analysis:

The identification and characterization of coding regions was carried out as follows: First, repetitive sequences were masked using RepeatMasker (A.F.A. Smit & P. Green, [http://ftp.genome.washington.edu/RM/RM\\_details.html](http://ftp.genome.washington.edu/RM/RM_details.html)) which screens DNA sequences in FastA format against a library of repetitive elements and returns a masked query sequence. Repeats not masked were identified by comparing the sequence to the GenBank database using WUBLAST2.0 [Altschul, S & Gish, W., Methods Enzymol. 266: 460-480 (1996); <http://blast.wustl.edu/blast/README.html>] and were masked manually.

Next, known genes were revealed by comparing the genomic regions against Genentech's protein database using the WUBLAST2.0 algorithm and then annotated by aligning the genomic and cDNA sequences for each gene, respectively, using a Needleman-Wunch (Needleman and Wunsch, J. Mol. Biol. 48: 443-453 (1970) algorithm to find regions of local identity between sequences. The strategy results in detection of all exons of the five known genes in the region, THPO, TRAP2, eIF4g, CLCN2 and hRPB17 (see below).

#### Known genes

eukaryotic translation initiation factor 4 gamma  
thrombopoietin  
chloride channel 2  
TNF receptor associated protein 2  
RNA polymerase II subunit hRPB17

#### Map position

3q27-qter  
3q26-q27  
3q26-qter  
not previously mapped  
not previously mapped

Finally, novel transcription units were predicted using a number of approaches. CpG islands (S. Cross & Bird, A., Curr. Opin. Genet. Dev. 5: 109-314 (1995) islands were used to define promoter regions and were

identified as clusters of sites cleaved by enzymes recognizing GC-rich, 6 or 8-mer palindromic sequences (NotI, NarI, BssHII, XhoI. CpG islands are usually associated with promoter regions of genes. WUBLAST2.0 analysis of short genomic regions (10-20 kb) versus GenBank revealed matches to ESTs. The individual EST sequences (or where possible, their sequence chromatogram files) were retrieved and assembled with Sequencer to provide a theoretical cDNA sequence (DNA36443). GRAIL2 (ApoCom Inc., Knoxville, TN, command line version for the DEC alpha) was used to predict a novel exon. The five known genes in the region served as internal controls for the success of the GRAIL algorithm.

#### Isolation:

A partial endothelin converting enzyme-2 (ECE-2) cDNA clone was isolated by first splicing in silico the ECE-2 exons predicted in the genomic sequence to generate a putative sequence (DNA36443). An oligonucleotide probe: GAAGCAGTGCAGCCAGCAGTAGAGAGGCACCTGCTAAGA (SEQ ID NO:530) was designed and used to screen a human fetal small intestine library (LIB110) and internal PCR primers (36443f1) (ECE2.f:ACGCAGCTGGAGCTGGTCTTAGCA) (SEQ ID NO:531) and (36443r1) (ECE2.r) (GGTACTGGACCCTAGGGCCACAA) (SEQ ID NO:532) were used to confirm clones hybridizing to the probe prior to sequencing. One positive clone was obtained, however this cDNA (DNA49830) represented a partially spliced transcript containing appropriately spliced exons 1 through 6, followed by intron 6 sequence. The oligo dT primer annealed to a polyA-stretch within an Alu element present in intron 6. An additional ECE-2 cDNA fragment (DNA49831) was obtained by PCR from a human fetal kidney library (LIB227) with primers designed from the presumed cDNA sequence [36443f3: CCTCCCAGCCGAGACCAGTGG (SEQ ID NO:533) and 36443r2: GGTCTATAAGGGCCAAGACC (SEQ ID NO:534)]. This PCR product extended from exon 13 into the 3' untranslated region in exon 18.

A full length endothelin converting enzyme 2 (ECE-2) cDNA clone (DNA55800-1263) was isolated from an oligo-dT-primed human fetal brain library. RNA from human fetal brain tissue (20 weeks gestation, #283005)(SRC175) was isolated by guanidine thiocyanate and 5 µg used to generate double stranded cDNA which was cloned into the vector pRK5E. The 3' -primer (pGACTAGTTCTAGATCGCGAGCGGCCGCCCTTTTTTTTTTTTTT) (SEQ ID NO:535) and the 5 -linker (pCGGACGCGTGGGTCGA) (SEQ ID NO:536) were designed to introduce XhoI and NotI restriction sites. The library was screened with PCR primers [36443pcrf1: CGGCCGTGATGGTGTGTGACG (SEQ ID NO:537) and 36443r3: GGCAGACTCCTTCCTATGGG (SEQ ID NO:538)] designed from the partial human ECE-2 cDNA sequences (DNA49830 and DNA49831). PCR products were cloned into the vector pCR2.1-TOPO (Invitrogen Corp., Carlsbad, CA, Cat. No. K4500-01) and sequenced with DYE-terminator chemistry as described above.

#### EXAMPLE 98: Northern Blot and in situ RNA Hybridization Analysis for PRO403

Expression of PRO403 mRNA in human tissues was examined by Northern blot analysis. Human polyA+ RNA blots derived from human fetal and adult tissues (Clontech, Palo Alto, CA; Cat. Nos. 7760-1, 7756-1 and 7755-1) were hybridized to a [32P-α]dATP-labelled cDNA fragments from probe based on the full length PRO403 cDNA. Blots were incubated with the probes in hybridization buffer (5X SSPE; 2X Denhardt's solution; 100 mg/mL denatured sheared salmon sperm DNA; 50% formamide; 2% SDS) for 18 hours at 42°C,

washed to high stringency (0.1XSSC, 0.1% SDS, 50°C) and autoradiographed. The blots were developed after overnight exposure by phosphorimager analysis (Fuji).

PRO403 mRNA transcripts were detected. Analysis of the expression pattern showed the strongest signal of the expected 3.3 kb transcript in adult brain (highest in the cerebellum, putamen, medulla, and temporal lobe, and lower in the cerebral cortex, occipital lobe and frontal lobe), spinal cord, lung and pancreas and higher levels of a 4.5 kb transcript in fetal brain and kidney.

#### EXAMPLE 99: Use of PRO Polypeptide-Encoding Nucleic Acid as Hybridization Probes

The following method describes use of a nucleotide sequence encoding a PRO polypeptide as a hybridization probe.

DNA comprising the coding sequence of a PRO polypeptide of interest as disclosed herein may be employed as a probe or used as a basis from which to prepare probes to screen for homologous DNAs (such as those encoding naturally-occurring variants of the PRO polypeptide) in human tissue cDNA libraries or human tissue genomic libraries.

Hybridization and washing of filters containing either library DNAs is performed under the following high stringency conditions. Hybridization of radiolabeled PRO polypeptide-encoding nucleic acid-derived probe to the filters is performed in a solution of 50% formamide, 5x SSC, 0.1% SDS, 0.1% sodium pyrophosphate, 50 mM sodium phosphate, pH 6.8, 2x Denhardt's solution, and 10% dextran sulfate at 42°C for 20 hours. Washing of the filters is performed in an aqueous solution of 0.1x SSC and 0.1% SDS at 42°C.

DNAs having a desired sequence identity with the DNA encoding full-length native sequence PRO polypeptide can then be identified using standard techniques known in the art.

#### EXAMPLE 100: Expression of PRO Polypeptides in *E. coli*

This example illustrates preparation of an unglycosylated form of a desired PRO polypeptide by recombinant expression in *E. coli*.

The DNA sequence encoding the desired PRO polypeptide is initially amplified using selected PCR primers. The primers should contain restriction enzyme sites which correspond to the restriction enzyme sites on the selected expression vector. A variety of expression vectors may be employed. An example of a suitable vector is pBR322 (derived from *E. coli*; see Bolivar et al., Gene, 2:95 (1977)) which contains genes for ampicillin and tetracycline resistance. The vector is digested with restriction enzyme and dephosphorylated. The PCR amplified sequences are then ligated into the vector. The vector will preferably include sequences which encode for an antibiotic resistance gene, a trp promoter, a polyhis leader (including the first six STII codons, polyhis sequence, and enterokinase cleavage site), the specific PRO polypeptide coding region, lambda transcriptional terminator, and an argU gene.

The ligation mixture is then used to transform a selected *E. coli* strain using the methods described in Sambrook et al., supra. Transformants are identified by their ability to grow on LB plates and antibiotic resistant colonies are then selected. Plasmid DNA can be isolated and confirmed by restriction analysis and DNA sequencing.

Selected clones can be grown overnight in liquid culture medium such as LB broth supplemented with antibiotics. The overnight culture may subsequently be used to inoculate a larger scale culture. The cells are then grown to a desired optical density, during which the expression promoter is turned on.

After culturing the cells for several more hours, the cells can be harvested by centrifugation. The cell pellet obtained by the centrifugation can be solubilized using various agents known in the art, and the solubilized PRO polypeptide can then be purified using a metal chelating column under conditions that allow tight binding of the protein.

PRO181, PRO195, PRO200, PRO237, PRO273, PRO540, PRO322, PRO1017, PRO938, PRO162, PRO1114, PRO827 and PRO1008 were expressed in *E. coli* in a poly-His tagged form, using the following procedure. The DNA encoding the PRO polypeptide was initially amplified using selected PCR primers. The primers contained restriction enzyme sites which correspond to the restriction enzyme sites on the selected expression vector, and other useful sequences providing for efficient and reliable translation initiation, rapid purification on a metal chelation column, and proteolytic removal with enterokinase. The PCR-amplified, poly-His tagged sequences were then ligated into an expression vector, which was used to transform an *E. coli* host based on strain 52 (W3110 fuhA(tonA) lon galE rpoHts(htpRts) clpP(lacIq)). Transformants were first grown in LB containing 50 mg/ml carbenicillin at 30°C with shaking until an O.D.600 of 3-5 was reached. Cultures were then diluted 50-100 fold into CRAP media (prepared by mixing 3.57 g (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 0.71 g sodium citrate-2H<sub>2</sub>O, 1.07 g KCl, 5.36 g Difco yeast extract, 5.36 g Sheffield hycase SF in 500 mL water, as well as 110 mM MPOS, pH 7.3, 0.55% (w/v) glucose and 7 mM MgSO<sub>4</sub>) and grown for approximately 20-30 hours at 30°C with shaking. Samples were removed to verify expression by SDS-PAGE analysis, and the bulk culture is centrifuged to pellet the cells. Cell pellets were frozen until purification and refolding.

*E. coli* paste from 0.5 to 1 L fermentations (6-10 g pellets) was resuspended in 10 volumes (w/v) in 7 M guanidine, 20 mM Tris, pH 8 buffer. Solid sodium sulfite and sodium tetrathionate is added to make final concentrations of 0.1M and 0.02 M, respectively, and the solution was stirred overnight at 4°C. This step results in a denatured protein with all cysteine residues blocked by sulfitolization. The solution was centrifuged at 40,000 rpm in a Beckman Ultracentrifuge for 30 min. The supernatant was diluted with 3-5 volumes of metal chelate column buffer (6 M guanidine, 20 mM Tris, pH 7.4) and filtered through 0.22 micron filters to clarify. Depending the clarified extract was loaded onto a 5 ml Qiagen Ni-NTA metal chelate column equilibrated in the metal chelate column buffer. The column was washed with additional buffer containing 50 mM imidazole (Calbiochem, Utrol grade), pH 7.4. The protein was eluted with buffer containing 250 mM imidazole. Fractions containing the desired protein were pooled and stored at 4°C. Protein concentration was estimated by its absorbance at 280 nm using the calculated extinction coefficient based on its amino acid sequence.

The proteins were refolded by diluting sample slowly into freshly prepared refolding buffer consisting of: 20 mM Tris, pH 8.6, 0.3 M NaCl, 2.5 M urea, 5 mM cysteine, 20 mM glycine and 1 mM EDTA. Refolding volumes were chosen so that the final protein concentration was between 50 to 100 micrograms/mL. The refolding solution was stirred gently at 4°C for 12-36 hours. The refolding reaction was quenched by the addition of TFA to a final concentration of 0.4% (pH of approximately 3). Before further purification of the protein, the solution was filtered through a 0.22 micron filter and acetonitrile was added to 2-10% final concentration. The refolded protein was chromatographed on a Poros R1/H reversed phase column using a